GENETICS

Clustering of vomeronasal receptor genes is required for transcriptional stability but not for choice

Quentin Dietschi[1]†, Joël Tuberosa[1]†, Leon Fodoulian[1,2]†, Madlaina Boillat[1], Chenda Kan[1], Julien Codourey[1], Véronique Pauli[1], Paul Feinstein[3], Alan Carleton[2]*‡, Ivan Rodriguez[1]*‡

Rodents perceive pheromones via vomeronasal receptors encoded by highly evolutionarily dynamic Vr and Fpr gene superfamilies. We report here that high numbers of V1r pseudogenes are scattered in mammalian genomes, contrasting with the clustered organization of functional V1r and Fpr genes. We also found that V1r pseudogenes are more likely to be expressed when located in a functional V1r gene cluster than when isolated. To explore the potential regulatory role played by the association of functional vomeronasal receptor genes with their clusters, we dissociated the mouse *Fpr-rs3* from its native cluster via transgenesis. Singular and specific transgenic *Fpr-rs3* transcription was observed in young vomeronasal neurons but was only transient. Our study of natural and artificial dispersed gene duplications uncovers the existence of transcription-stabilizing elements not coupled to vomeronasal gene units but rather associated with vomeronasal gene clusters and thus explains the evolutionary conserved clustered organization of functional vomeronasal genes.

INTRODUCTION

In mammals, vomeronasal perception plays a major role in interindividual interactions, including predator detection, reproduction, and avoidance of sick conspecifics (*1*, *2*). It is thus critical not only for the survival of individuals but also to the maintenance of their corresponding species.

The vomeronasal organ (VNO), an elongated olfactory structure located at the base of the nasal cavity of most mammals, contains sensory neurons that express receptors specialized in the detection of pheromones and kairomones (*3*). These chemoreceptors are seven transmembrane G-coupled receptors, the vomeronasal type 1 (V1r), type 2 (V2r), or formyl peptide (Fpr) receptor families (*4*). V1r gene repertoires are among the most variable in size among vertebrate gene families. For example, the platypus, the mouse, and the elephant genomes contain 302, 235, and 34 V1r genes, respectively (*5*). For example, entire V1r subfamilies are often specific to a given species, even between closely related species such as mice and rats. This diversity is the result of an exceptionally dynamic evolutionary history that involved numerous and rapid gene births and deaths.

Fprs, which are expressed by immune cells in mammals, have acquired olfactory-specific expression in rodents (*6*, *7*). The neuronal specificity of *Fpr-rs3*, *rs6*, *rs7*, and *rs4* was acquired after the hijack of a V1r promoter (whose sequence is still conserved today) by a duplicated immune Fpr coding sequence (CDS) about 30 million years ago, followed by the multiplication and divergence of this chimeric gene (*8*). As a result, these four Fprs are embedded inside a vomeronasal receptor gene cluster. Similar to V1rs (*9*, *10*), they are expressed in the apical part of the vomeronasal epithelium in a punctate and monogenic pattern (*6*, *7*), and similar to V1rs, their corresponding sensory neurons project to the rostral accessory olfactory bulb (AOB) (*11*). Given their evolutionary history, expression patterns, and promoter conservation, *Fpr-rs3*, *rs6*, *rs7*, and *rs4* transcriptional regulation is considered similar to that of the V1rs (*8*).

We report here that, in contrast to vomeronasal pseudogenes that are scattered throughout the genome, functional vomeronasal receptor genes exhibit a conserved clustered organization. Using a combination of genomic, transcriptomic and transgenic approaches, we investigated the forces that may have been at work in the maintenance and expansion of vomeronasal receptor gene clusters during evolution. Our data suggest that it is not vomeronasal receptor gene choice but rather the continuation of transcription that represents the main selective force that maintains the clustering of functional vomeronasal receptor genes.

RESULTS

Different genomic organization of V1r/Fpr pseudogenes and functional genes

Among mammals, the diversity of vomeronasal receptor genes reflects their fast evolution. To potentially expand this observation and to put this rapid evolution in context, we compared the vomeronasal receptor diversity with that of odorant receptors (Ors), another rapidly evolving and large chemoreceptor superfamily. We extracted the complete V1r and Or gene repertoires from the genomes of 29 mammalian species and found 1724 and 22,885 genes, respectively. To cover most major lineages, we analyzed genomes of Glires, Primates, Laurasiatheria, Xenarthra, Marsupialia, Afrotheria, and Monotremata. To investigate the evolution dynamics of the repertoires, we built a phylogeny from the Or and V1r protein sequences of all species. A pattern emerged, pointing to a remarkable proportion of V1r species-specific paralogs (inparalogs, i.e., from gene duplications that occurred after the last speciation event included in our analysis), more pronounced than for Ors (Fig. 1A). To evaluate the significance of this difference, we identified all V1r and Or inparalog groups of the 29 species, classified them by size, and compared their contribution to their corresponding repertoires. We found that V1r inparalog groups (that means monophyletic and species-specific paralogs) containing four or more genes represent a larger proportion of each species

[1]Department of Genetics and Evolution, Faculty of Sciences, University of Geneva, Geneva, Switzerland. [2]Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland. [3]Department of Biological Sciences, Hunter College, City University of New York and The Graduate Center Programs in Biochemistry, Biology and CUNY Neuroscience Collaborative, New York, NY, USA.
*Corresponding author. Email: alan.carleton@unige.ch (A.C.); ivan.rodriguez@unige.ch (I.R.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

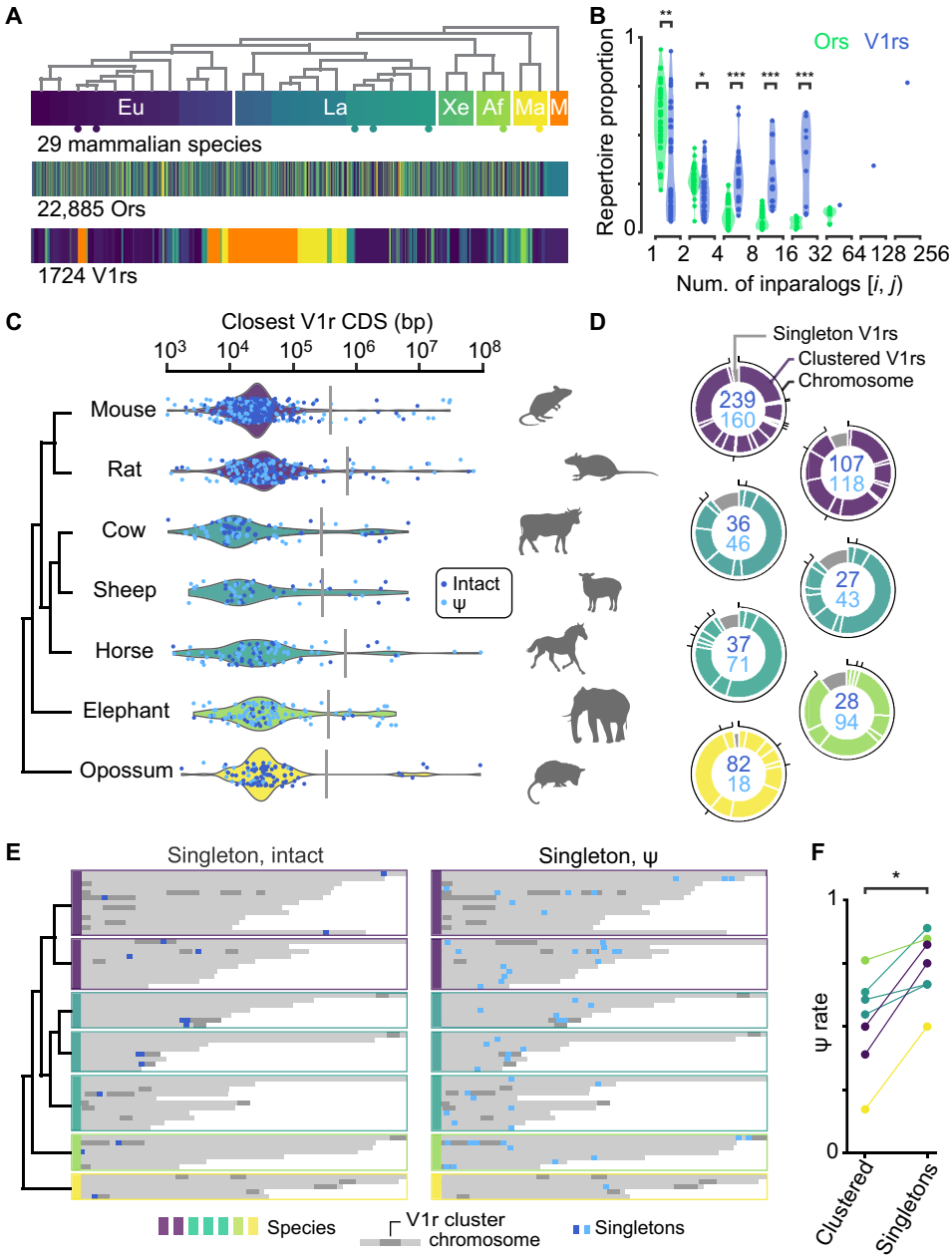

Fig. 1. Different genomic organization of V1r/Fpr pseudo- and functional genes. (**A**) Phylogeny of 29 mammals with major clades labeled: Eu-archontoglires, La-urasiatheria, Xe-narthra, Af-rotheria, Ma-rsupialia, and M-onotremata. Dots indicate the species analyzed in (C) to (F). Or and V1r genes as tiled colored by species and ordered as they appear on their respective phylogenies are shown below. (**B**) Gene duplication rate within Or (green) and V1r (blue) repertoires, evaluated through the size of species-specific paralog (inparalog) groups (*x* axis labels represent size bins such as *i* <= inparalog number < *j*). Left to right: $^{*}P = 0.032$, $^{**}P = 0.0029$, $^{***}P = 1.3 \times 10^{-7}$, $^{***}P = 5.7 \times 10^{-5}$, and $^{***}P = 2.2 \times 10^{-5}$, Wilcoxon. Density estimates are drawn in the background. (**C**) Distances between V1r genes. Left: Species phylogeny. Dot plots: Distance between adjacent intact V1r CDSs (dark blue) or V1r pseudogenes (ψ, light blue). Density estimates are drawn in the background. Gray vertical bars: clustering thresholds. (**D**) V1r gene and cluster numbers. Colored circle segments show the relative size of the different gene clusters. Gray segment represents the proportion of singletons. Outer edge segments delimit chromosomes or assembly scaffolds. Total numbers of intact V1r genes (dark blue) and pseudogenes (light blue) are displayed. (**E**) Genomic distribution of singleton V1r genes within the genome of species featured in (C) and (D). For each species, chromosomes or scaffolds that contain at least one V1r sequence are shown as horizontal gray bars of proportional lengths. (**F**) Clustered versus singleton V1r pseudogene (Ψ) rates. $^{*}P = 0.016$, Wilcoxon.

repertoire than Or inparalog groups of the same size (Fig. 1B). The evolution of species-specific mammalian V1r gene repertoires via gene duplication is thus faster than the one of Or genes.

Mouse V1r and Fpr genes are organized in clusters. To evaluate how generalizable to other mammalian lineages this observation may be, we analyzed the genomic arrangements of all V1r and vomeronasal Fpr genes (including functional genes and pseudogenes) present in the genomes of the mouse (*Mus musculus*), the rat (*Rattus norvegicus*), the cow (*Bos taurus*), the sheep (*Ovis aries*), the elephant (*Loxodonta africana*), and the opossum (*Monodelphis domestica*). To identify gene

cluster boundaries, we defined an aggregation threshold based on the distribution of intergenic distances measured between neighboring and consecutive V1r sequences on each chromosome (intact or pseudogenic CDSs) (Fig. 1C). We found that in all species analyzed, V1r and vomeronasal Fpr genes were organized in clusters, with a relatively conserved distance between members of a given cluster among the different species (Fig. 1D). The number of clusters ranged from 8 in the cow to 18 in the mouse (a mean of 10.8), with a size reaching 105 members in the mouse (a mean of 37) (Fig. 1D). Genes that could not be clustered were considered as singletons and represented a minor portion of the repertoire in each species (Fig. 1D).

Strikingly, we found that most of the singletons were pseudogenes (Fig. 1E). To evaluate the survival rate of singleton genes outside clusters, we compared the rate of pseudogenes within singleton V1rs with the rate of pseudogenes within clustered V1rs. We found a significantly higher proportion of pseudogenes among singletons ($P = 0.016$; Fig. 1F).

These data indicate (i) that, despite a dynamic history of V1r/Fpr gene birth and death, functional V1r and Fpr genes remained clustered during evolution and (ii) that a high proportion of singletons are pseudogenes. Together, these findings suggest that to retain function, V1r and Fpr genes must remain clustered.

Transcriptional profiles of different V1r genes and clusters

Each V1r/Fpr gene is thought to be randomly chosen for transcription. Our observations relative to V1r clusters and to V1r pseudogenes outside clusters suggest that regulatory elements, possibly shared by multiple V1rs within a given cluster, may affect aspects of V1r transcription. To evaluate whether some of these aspects, particularly the frequency at which a given V1r is chosen by a vomeronasal sensory neuron (VSN), could be biased toward specific V1rs and toward those pertaining to specific clusters, we evaluated the number of transcripts corresponding to each V1r and Fpr genes. To this aim, we sequenced the transcriptome of adult mouse VNOs (Fig. 2A). We observed cluster-specific levels of V1r transcripts. For example, clusters 3 and 6 (on chromosome 7) were significantly less transcribed than cluster 5 (on chromosome 6) (Fig. 2B). To determine whether this observation reflects differences in probability of choice or differences in transcription levels across V1r genes in different clusters, we isolated VSNs and performed single-cell RNA sequencing (scRNA-seq; Fig. 2A). We observed that vomeronasal receptor genes pertaining to some clusters were more likely to be chosen for expression than vomeronasal receptor genes pertaining to other clusters [compare the number of neurons expressing specific members of the cluster 5 (chromosome 6) relative to those from the cluster 6 (chromosome 7)] (Fig. 2C). The number of vomeronasal neurons obtained in the single-cell analysis and expressing a given V1r was positively correlated with the amount of corresponding mRNAs in the whole-tissue RNA-seq analysis (Fig. 2D). Thus, the probability of an intact V1r to be transcribed appears dependent, at least in part, on the cluster in which it is located. Last, whereas most of mouse singleton V1rs are pseudogenes (9 of 12), three functional singletons are expressed at substantial levels (Fig. 2B), representing rare cases where conditions for transcript stabilization are met outside of a V1r/Fpr gene cluster.

We then reasoned that if cluster-specific regulatory elements affect the transcription of functional V1r genes, the same elements may also affect V1r pseudogene transcription. We first evaluated the potential transcriptional activity of V1r pseudogenes and observed, as expected, very low levels of V1r pseudogene transcription relative to the ones characterizing intact V1r genes (Fig. 2, B and E). We then compared the transcription of the V1r pseudogenes located in proximity to gene clusters containing functional V1rs/Fprs relative to those isolated from these clusters. Given the limited number of expressed V1r pseudogenes in the mouse (Fig. 2B), this analysis was performed on both mouse and rat vomeronasal transcriptomes. We observed a significant difference in the level of expression between rat pseudogenes that are associated with V1r/Fpr gene clusters and those that are not (Fig. 2F). This further suggests an association between cluster proximity and transcriptional activity.

Fpr-rs3 expression is not recapitulated when placed outside its endogenous genomic context

The unequal probabilities of V1rs pertaining to specific clusters to be transcribed may result from cluster-specific regulatory elements that are either physically shared between V1r genes or that are associated with each gene. In this latter case, a defined probability for promoters pertaining to a cluster to be active in mature neurons would be explained by the fact that V1r genes pertaining to a given cluster usually belong to a unique V1r subfamily and therefore share not only their CDSs but also characteristic promoters. Alternatively, shared transcriptional regulatory elements present in V1rs/Fprs clusters may act on a few V1rs/Fprs or even on all members of a cluster and provide V1rs and Fprs a cluster-specific probability to be selected for expression. The proximity of a gene to a given V1r cluster regulatory element(s) could here play a role in transcriptional characteristics. Our data on the functionality of V1r/Fpr singletons (intact versus pseudogenic), relative to the transcriptional levels of clustered versus singleton pseudogenes, point to this second alternative.

To further explore the potential role played by V1r/Fpr gene clustering, we evaluated the expression of a vomeronasal chemoreceptor gene when located in its endogenous gene cluster or when experimentally isolated from its natural cluster. To this aim, we modified the genomic landscape surrounding a specific Fpr/V1r gene by integrating a specific Fpr/V1r gene away from its endogenous cluster. To do this, we chose *Fpr-rs3* (Fig. 3A), a highly transcribed Fpr vomeronasal receptor gene that we previously showed resulted from a transposition in the rodent lineage and hijacked a V1r promoter, conserving both the promoter sequence and tissue specificity features of the original V1r. We first generated a knockin line (*Fpr-rs3*^{cre}), in which a polycistronic cassette was added to the endogenous *Fpr-rs3*, such that it led to the cotranscription and cotranslation of a fluorophore and of the Cre recombinase (Fig. 3B). In parallel, we generated two different *Fpr-rs3*–expressing transgenes. One was driven by a highly conserved 400–base pair (bp) segment upstream of the *Fpr-rs3* transcriptional start [a sequence conserved among vomeronasal Fpr genes and among the corresponding V1r genes in nonrodent mammalian species (*8*)]. The second transgene was very large and included 142 kb surrounding the *Fpr-rs3* CDS (Fig. 3B). The CDS of a fluorophore and of the Cre recombinase was placed, similar to the knockin line, as polycistrons, after the *Fpr-rs3* stop codon. Three transgenic lines [Tg(Fpr-rs3-cre)2, Tg(Fpr-rs3-cre)5, and Tg(Fpr-rs3-cre)7] and two transgenic lines [Tg(BAC-Fpr-rs3-cre)2 and Tg(BAC-Fpr-rs3-cre)4] corresponding to the first and second transgenes were generated, respectively. Coronal vomeronasal sections of 6-week-old *Fpr-rs3*$^{cre/+}$, Tg(Fpr-rs3-cre)2/5/7, and Tg(BAC-Fpr-rs3-cre)2/4 mice were analyzed to evaluate *Fpr-rs3* expression (Fig. 3, C to G).

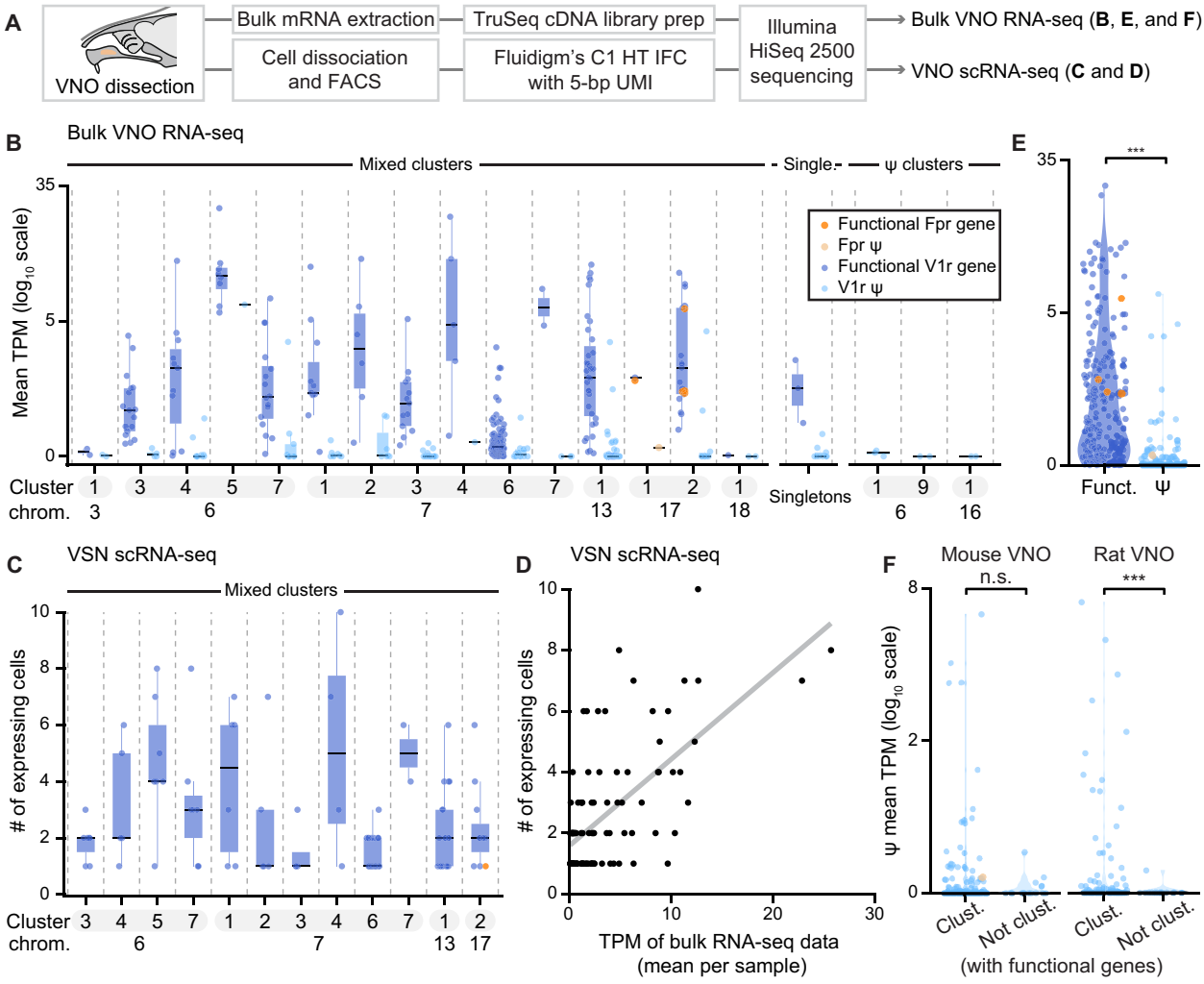

Fig. 2. Transcriptional profiles of different V1r genes, pseudogenes, and clusters. (**A**) Experimental paradigm. IFC, integrated fluidic circuit; FACS, fluorescence-activated cell sorter; HT, high throughput; bp, base pair; UMI, unique molecular identifier. (**B**) V1rs and apical Fprs transcription levels quantified by whole-tissue RNA-seq of adult mouse VNOs ($n = 8$). Dots represent mean transcripts per million (TPM) values of single genes arranged by gene clusters. Clusters are sorted in three categories: "mixed cluster" (with functional genes and pseudogenes), "single" (isolated genes), and "Ψ" (with pseudogenes only). Effect of the cluster identity on the expression level was tested for functional genes of the "mixed clusters" group [analysis of variance (ANOVA): $P < 2 \times 10^{-16}$]. wt, wild type. (**C**) scRNA-seq of VSNs, dots represent the numbers of cells expressing a given functional V1r or Fpr gene identified ($P = 7$ mice, 98 different VSN types). Effect of the cluster belonging on VSN abundance was tested with an ANOVA ($P = 0.000232$). (**D**) Correlation between the number of cells expressing a given V1r detected in the scRNA-seq dataset and the transcript abundance of this V1r in the bulk RNA-seq dataset (Kendall's rank correlation $\tau = 0.405$, $P = 6.967 \times 10^{-08}$). (**E**) Mean expression levels of functional genes (funct.) and pseudogenes (Ψ). Distribution shift between functional genes and pseudogenes was tested with a Wilcoxon ($W = 33822$, $P < 2.2 \times 10^{-16}$). (**F**) Pseudogene transcription levels within or outside mixed clusters. Left: $n = 8$ mouse VNOs, 141 Ψ in mixed clusters, and 16 Ψ outside. Right: $n = 7$ rat VNOs, 102 Ψ in mixed clusters, and 16 Ψ outside. Distribution shifts between the two categories were tested with Wilcoxon (mouse: $W = 1214$, $P = 0.5704$; rat: $W = 1113$, $P = 0.002987$). n.s., not significant.

Taking advantage of the green fluorophore coexpressed with *Fpr-rs3* in all three transgenic versions, the number of fluorescent VSNs was evaluated. The *Fpr-rs3* knockin line showed a mean of 190 fluorescent neurons per VNO (Fig. 3, D and G), in line with the number of *Fpr-rs3*–expressing neurons we previously estimated by in situ hybridization and immunohistochemistry (*6*, *11*). Unexpectedly, no or very few VSNs expressed the transgene in any of the three Tg(Fpr-rs3-cre) lines nor even in the two Tg(BAC-Fpr-rs3-cre) lines (Fig. 3, E to G). Insertion of *Fpr-rs3* transgenes outside the endogenous *Fpr-rs3*–containing cluster, even when containing substantial amounts of genomic information surrounding the endogenous *Fpr-rs3*, does thus not recapitulate the endogenous *Fpr-rs3* expression pattern.

Outside the V1r/Fpr cluster, *Fpr-rs3* is chosen but not stabilized

The nonrecapitulation of *Fpr-rs3* expression by the transgenes may result from different deficiencies. Either the transgenes lack regulatory elements necessary for proper choice and expression of *Fpr-rs3*, or, alternatively, the transgenes may be selected for expression, but this transcription might be transient if the neuron dies or if the choice is not stabilized. To test these different hypotheses, we took advantage of the Cre recombinase expressed by the transgenes and crossed the Tg(Fpr-rs3-cre)2/5/7 and Tg(BAC-Fpr-rs3-cre)2/4 transgenic lines with a Cre-dependent reporter strain, *Rosa*^{stopRFP}, which drives expression of a red fluorophore after Cre-mediated recombination (Fig. 4A). No expression of the reporter was observed outside the

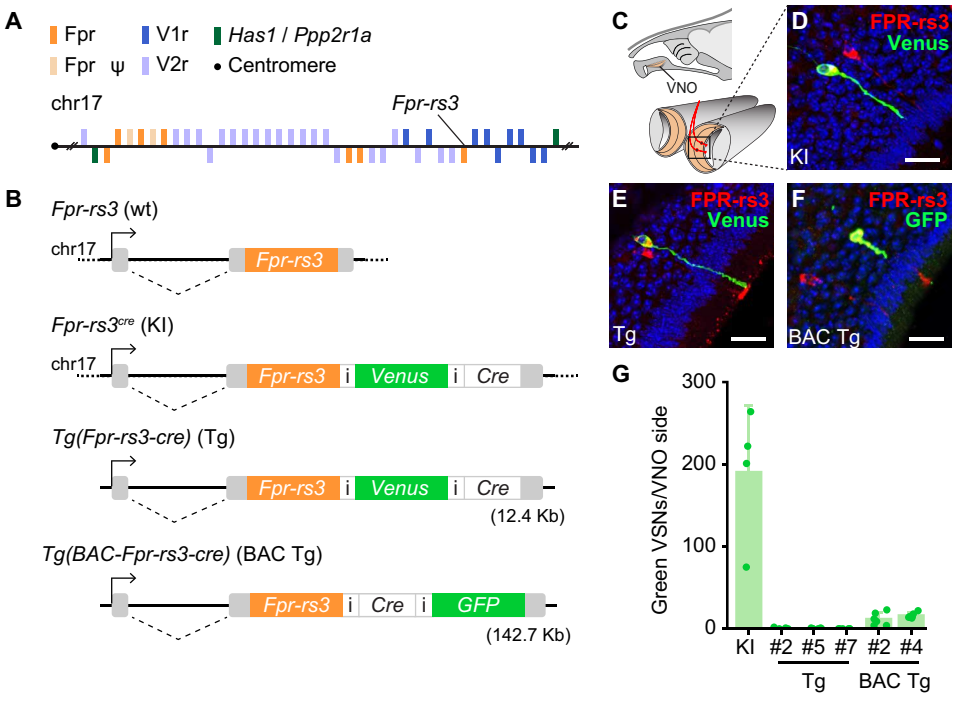

Fig. 3. *Fpr-rs3* expression is not recapitulated when placed outside its endogenous genomic context. (**A**) Topology of the Fpr/Vr gene cluster on chromosome 17. Colored boxes represent genes along the chromosome 17, with the centromere on the left (assembly: GRCm38/mm10). Genes are displayed above or below the chromosome line according to their transcriptional direction (above: same as reference; below: reverse). (**B**) *Fpr-rs3* transgenes. KI, knockin; Tg, short transgene with short intron; BAC Tg, bacterial artificial chromosome transgene (142 kb). (**C**) Schematic of a mouse head hemisection and of the two sides of the VNO (sectioned coronally). *Fpr-rs3* VSNs (red) are situated in the apical epithelium along with V1r neurons. (**D** to **F**) Double immunolabeling of Venus or green fluorescent protein (GFP) and FPR-rs3. Scale bars, 20 μm. (**G**) Number of cells per VNO side expressing the transgene (reflected by the number of green fluorescent VSNs). Bars and whiskers represent means and SDs, respectively. Two to three different P50 animals were analyzed per condition.

VNO. In contrast, a large number of neurons expressing the reporter was observed in the vomeronasal neuroepithelium of mice bearing any of the five transgenes [from 500 to 1429 red fluorescent neurons per VNO (Fig. 4, A to C) to be compared with 0 to 23 green fluorescent neurons, that is between 0 to 1.69% of red cells that are green]. In *Fpr-rs3*$^{cre/+}$; *Rosa*$^{stopRFP/+}$ knockin mice, the *Fpr-rs3* allele was stabilized in about 15% of the sensory neurons that initiated its transcription, suggesting an endogenous gene expression process that involves a substantial probability for a chemoreceptor gene of not being stabilized after being chosen, possibly switching to other chemoreceptor genes. In the corresponding short and long *Fpr-rs3* transgenes located outside of the endogenous cluster, the *Fpr-rs3* allele was stabilized in less than 2% of the neurons (Fig. 4C). An obvious explanation for these observations is that, when removed from its cluster, *Fpr-rs3* is subjected to choice during neuron maturation, and transcription is not stabilized.

Transgenic Fpr-rs3 expression is limited to young neurons

An alternative explanation of the silencing/switch phenomenon could however be proposed. One could consider the transient expression of *Fpr-rs3* as resulting from leaky transcription, in other words, from an unregulated and low level transcription of the transgenes, that would be sufficient to trigger recombination of the reporter. This noise would translate into red fluorescent neurons that are not green fluorescent. The transiently green neurons observed in transgenic mice would correspond to this noise. We tested this alternative explanation by evaluating multiple characteristics that should be observed in the unstabilized choice model, but that should be absent in the leaky transcription model.

Our current understanding of mammalian chemoreceptor gene choice involves the choice of a single receptor gene early during the life of a sensory neuron, followed by a maintenance of this choice during the life of this neuron. Thus, in the unstabilized choice model, one would expect only recently born neurons to show active transgene expression. In the leaky transcription model on the contrary, this temporal restriction would not necessarily be expected (Fig. 5). To evaluate the stage at which transgenic *Fpr-rs3* expression was taking place, we took advantage of the unequal spatial distribution of young and mature neurons in the vomeronasal neuroepithelium. This segregation can easily be visualized in coronal sections of the VNO, given that neurons are born in the vomeronasal crescent tips and migrate toward the center while maturing. We divided each vomeronasal crescent of *Fpr-rs3*$^{cre/+}$; *Rosa*$^{stopRFP/+}$ and Tg(BAC-Fpr-rs3-cre)2/4; *Rosa*$^{stopRFP/+}$ mice into three areas, two "immature" (dorsal and ventral) and one "mature" (medial) area, and recorded the number of neurons expressing the recombined reporter (red) and the transgene-driven fluorophore (green) (Fig. 5, A and C). We observed a very similar distribution of neurons expressing the red fluorescent protein (RFP) reporter in *Fpr-rs3*$^{cre/+}$; *Rosa*$^{stopRFP/+}$ and Tg(BAC-Fpr-rs3-cre)2/4; *Rosa*$^{stopRFP/+}$ mice (Fig. 5, A and B). However, when the distribution of neurons expressing the *Fpr-rs3*–driven YFP fluorophore was compared between the transgenic lines (n = 606 vomeronasal sections combined) and the knockin line, a marked reduction of green fluorescent neurons was observed in the mature

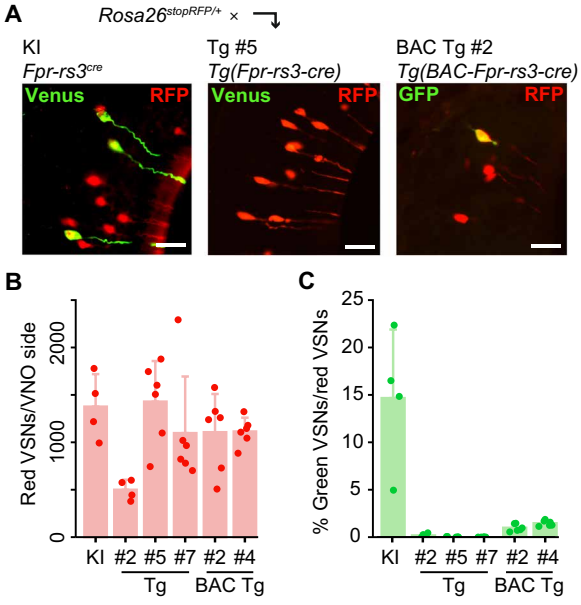

Fig. 4. Transient transcription of the *Fpr-rs3* transgenes located outside the V1r/Fpr cluster. (**A**) Mice bearing one of the *Fpr-rs3* transgenic alleles were crossed to *Rosa*$^{stopRFP/+}$ mice to permanently label the neurons that transcribed the transgene, even transiently. Representative epifluorescence microscopy images that were used to identify Cre-dependent reporter-expressing cells [with red fluorescent protein (RFP), in red] and transgene-expressing cells (with Venus or GFP, in green) in coronal sections of the VNO (schematic on the left). Scale bars, 20 μm (**B**) Numbers of RFP-expressing cells (referred to as red cells) per VNO side, measured in animals bearing one of the transgenic alleles (*x* axis). Dots: Single measures. Bars: Means and SEM. (**C**) Quantification of transgene-expressing cells (referred to as green cells) per red cells; same data are shown in (B).

zone of the transgenic animals (Fig. 5, C and D), in addition to the lower proportion of stabilized transgenic expression across the whole VNO (Fig. 5E). In parallel to this approach that used topographical localization as a proxy to neuronal age, we compared the percentage of neurons bearing a recombined reporter allele and still expressing the transgene between young (P7 and P15) and older mice (P30) (Fig. 5F). The idea was that, given the naturally higher ratio of young/old vomeronasal neurons in perinatal mice relative to 30-day-old animals, the percentage of green fluorescent protein (GFP)/RFP double-positive neurons in transgenic mice would be expected to go down with age if an early gene choice was followed by a transcriptional shut down as the neuron matures. We analyzed P7, P15, and P30 *Fpr-rs3*$^{cre/+}$*; Rosa*$^{stopRFP/+}$ and Tg(BAC-Fpr-rs3-cre)2/4*; Rosa*$^{stopRFP/+}$ mice. We observed a decrease with the age of the transgenic mice bearing the randomly integrated transgene in the percentage of GFP/RFP double-positive neurons, which was not seen in knockin animals (Fig. 5F). Together, these two independent approaches show that transgene-driven *Fpr-rs3* expression is mostly limited to young vomeronasal neurons, neurons that then continue to live with a silenced *Fpr-rs3* transgene, again favoring the unstabilized choice explanation (Fig. 5G).

High levels of transgenic Fpr-rs3 expression

Once a V1r allele is chosen and stabilized, it reaches an unusually high level of transcription. In the unstabilized choice model—that is choice followed by silencing—one would thus expect most chosen V1r alleles to reach at some point a substantial level of transcription. On the contrary, in the leaky transcription model, although a low level transcription of the Cre recombinase is necessary to drive recombination of loxP sites, this level would not be expected to approach the level of the endogenous *Fpr-rs3*. We evaluated the transcriptional activity of both endogenous and transgenic *Fpr-rs3* in *Fpr-rs3*$^{cre/+}$ and Tg(BAC-Fpr-rs3-cre)2 mice by measuring the level of *Cre* expression using RNAscope single-molecule mRNA fluorescent in situ hybridization in individual neurons. We found the *Cre* probe fluorescence of neurons expressing the Fpr-rs3 transgene to reach the same high intensity as the one of neurons expressing the endogenous *Fpr-rs3* gene (Fig. 5, H and I), supporting the unstabilized choice model. Thus, together, our data strongly support a model in which the broad transient expression of the transgenes occurs via early, unstabilized gene choice.

Transgenic Fpr-rs3 expression is punctate, restricted to Gαi2-expressing neurons, and exclusively present in neurons targeting the rostral AOB

If Cre-dependent cell marking were due to normal but unstabilized gene choice, then this labeling should follow normal cell-type specificity for V1r/Fpr genes. Alternatively, if Cre-dependent cell marking was due to nonspecific, leaky transcription in all vomeronasal cell types, then we would expect Cre-dependent reporter expression in most vomeronasal neurons. This was not observed. In all lines analyzed, transgenic expression of *Fpr-rs3* was exclusively observed in vomeronasal neurons and appeared to be restricted to apically located (presumably Gαi2-expressing) vomeronasal neurons, as expected for V1r-specific choice. To examine this more closely, we characterized the identity of the neurons that transiently express *Fpr-rs3*. We first colabeled VSNs from P30 *Fpr-rs3*$^{cre/+}$*; Rosa*$^{stopRFP/+}$ and Tg(Fpr-rs3-cre)5*; Rosa*$^{stopRFP/+}$ mice with RFP and Gαi2 (a marker of V1r/Fpr-expressing neurons, absent in V2r positive vomeronasal neurons). More than 85% of the red fluorescent neurons were also positive for Gαi2 (Fig. 6, A to E, and fig. S1, A to C). We also colabeled VSNs of the same animals with RFP and Gαo (a marker of V2r-expressing neurons, absent in V1r-positive vomeronasal neurons). Over 90% of the red neurons were negative for Gαo (Fig. 6, F to H). Second, we investigated the axonal projections of *Fpr-rs3*$^{cre/+}$*; Rosa*$^{stopRFP/+}$, Tg(BAC-Fpr-rs3-cre)2/4*; Rosa*$^{stopRFP/+}$, and Tg(Fpr-rs3-cre)5*; Rosa*$^{stopRFP/+}$ mice. All red fluorescent fibers targeted the rostral accessory olfactory bulb (Fig. 6, I to P, and fig. S1D) (which corresponds to the V1r/Fpr targeting zone), while the caudal accessory olfactory bulb (the V2r targeting zone) was devoid of red axons. Which proportion of these red fluorescent axons innervating most areas of the rostral accessory bulb represent rerouted axons that did initially choose *Fpr-rs3* before switching receptor, or result from potentially leaky Cre expression in some mature neurons, remains to be determined. However, we indirectly explored the identity of these red fluorescent vomeronasal neurons by exploring their potential transcription of V1r, Fpr, and V2r genes. We performed in situ hybridizations on vomeronasal sections of *Fpr-rs3*$^{cre/+}$*; Rosa*$^{stopRFP/+}$, Tg(BAC-Fpr-rs3-cre)2*; Rosa*$^{stopRFP/+}$, and Tg(Fpr-rs3-cre)5*; Rosa*$^{stopRFP/+}$ mice with probes covering the repertoires of the three receptor types (fig. S1, E and F). We found expression of V1r and Fpr genes in red fluorescent neurons, but not of V2r genes. Last, we looked at axonal projections of Fpr-rs3–expressing neurons. We observed, as expected, a coalescence of green fibers into glomeruli in the rostral accessory olfactory bulb of *Fpr-rs3*$^{cre/+}$*; Rosa*$^{stopRFP/+}$ mice (Fig. 6, K and L) and a

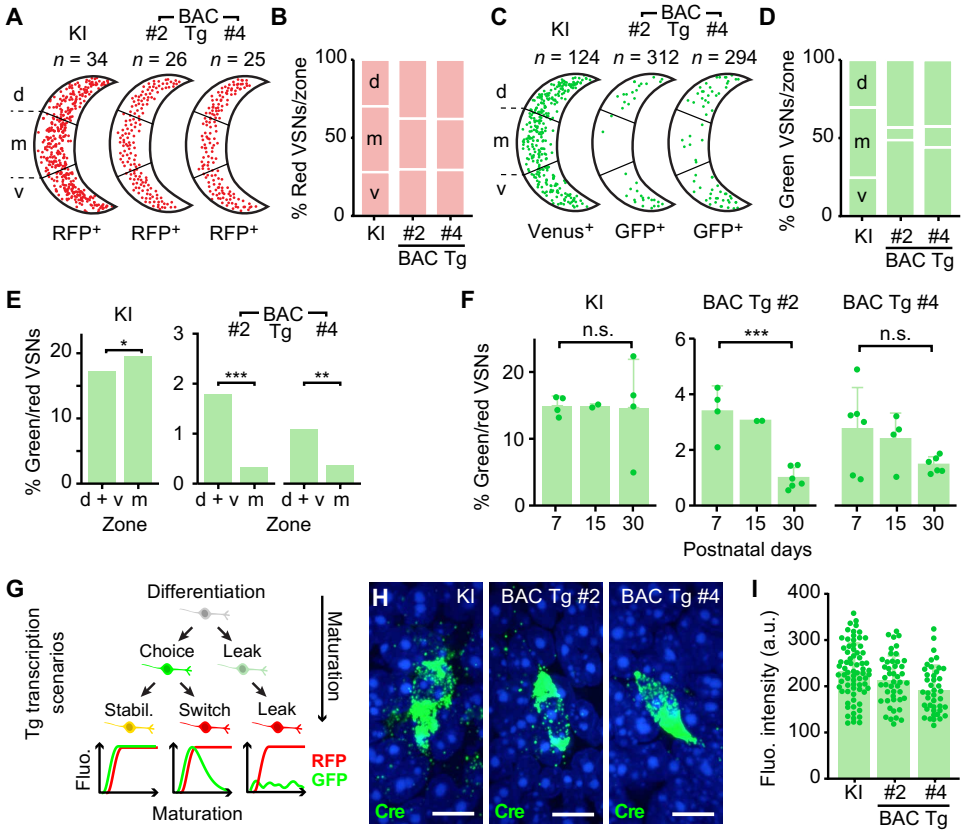

Fig. 5. A massive transcription of the transgenes takes place at the time of chemoreceptor gene choice but is not stabilized. (**A**) Stereotyped VNO crescent shapes of coronal sections were divided in three sectors (d, dorsal; m, medial; v, ventral). Each red dot represents an RFP$^+$ VSN in *Fpr-rs3*$^{cre/wt}$ (KI, $n = 2$) or Tg(BAC-Fpr-rs3-cre) lines [BAC Tg #2 ($n = 1$) or BAC Tg #4 ($n = 1$)]. (**B**) Stacked bars show the percentages of labeled cells represented in (A) in the different zones of the VNO. (**C** and **D**) Same sections but showing Venus$^+$ or GFP$^+$ VSNs, represented in green. (**E**) Rates of Venus$^+$ or GFP$^+$ cells [shown in (C)] among RFP$^+$ cells [shown in (A)] in the zones containing young (dorsal and ventral) or mature VSNs (medial). *$P = 0.0105$, **$P = 0.0026$, and ***$P = 6.67 \times 10^{-6}$, chi-squared. (**F**) Nonnormalized stabilized rate across the whole VNO and throughout animal maturation. n.s., $P = 0.686$ (KI); ***$P = 0.0095$; and n.s., $P = 0.393$ (BAC Tg #4), Wilcoxon rank sum test with continuity correction. (**G**) Different scenarios may account for the observed expression patterns. The canonical mechanism leading to monogenic expression includes choice and stabilization or choice and switch to another allele. In both cases, it involves the chosen receptor to reach a high level of expression, even transiently. In a leaky expression model, without expression burst, the constant and low level of Cre expression eventually leads to the accumulation of RFP$^+$ VSN, and green fluorescence never reaches comparable levels to that generated by the choice mechanism. (**H**) Representative images of VSNs from *Fpr-rs3*$^{cre/wt}$ and Tg(BAC-Fpr-rs3-cre) mouse lines labeled with a single-molecule mRNA fluorescent in situ hybridization probe against Cre (green) and counterstained with 4and Tgamidino-2-phenylindole (DAPI) (blue). Scale bars, 10 μm. (**I**) Quantification of mean fluorescence intensity of Cre probe. $n = 3$ mice were analyzed for each genotype, each dot corresponds to one cell (KI, $n = 72$; BAC Tg #2, $n = 46$; BAC Tg #2, $n = 42$). a.u., arbitrary units.

convergence to specific loci of the rare fibers that were still green fluorescent in Tg(BAC-Fpr-rs3-cre)2; *Rosa*$^{stopRFP/+}$ mice (Fig. 6, M and N, and fig. S1D). These observations show a cell-type specificity of transient *Fpr-rs3* expression in the V1r/Fpr vomeronasal subpopulation, consistent with choice not followed by its maintenance.

Transgenic Fpr-rs3 expression and apoptosis

An alternative to the choice followed by transcriptional shutdown would be a transient *Fpr-rs3* choice followed by death of the neurons that maintained transcription of the transgenic allele. We explored this possibility by evaluating potential apoptotic events in *Fpr-rs3*–expressing neurons during development. We identified GFP$^+$ neurons in coronal vomeronasal sections of P12 and P30 *Fpr-rs3*$^{cre/+}$; *Rosa*$^{stopRFP/+}$ and P12 Tg(Fpr-rs3-cre)5; *Rosa*$^{stopRFP/+}$ mice and tested their potential colabeling with the cell death marker Caspase 3. As a measure of the expected natural apoptotic rate in neurons at this developmental stage, we also quantified the apoptotic rate of vomeronasal neurons surrounding those expressing *Fpr-rs3*. We found a 1.31% rate of apoptosis in neurons surrounding green fluorescent neurons (35,765 neurons analyzed) and a 0% rate of apoptotic *Fpr-rs3*–expressing neurons in *Fpr-rs3*$^{cre/+}$; *Rosa*$^{stopRFP/+}$ and Tg(Fpr-rs3-cre)5; *Rosa*$^{stopRFP/+}$ mice (KI P12: $n = 159$; KI P30: $n = 237$; Tg #5 P12: $n = 164$ neurons analyzed). These data do not support an explanation involving a sustained transgenic *Fpr-rs3* transcription followed by apoptosis but rather a lack of transcriptional stabilization.

DISCUSSION

By initially taking a global genomic approach using multiple mammalian genomes (ranging from to rodents), we report that the genomic organization of pseudogenic and functional V1r/Fpr genes is different. We found that functional V1r/Fpr genes are organized in

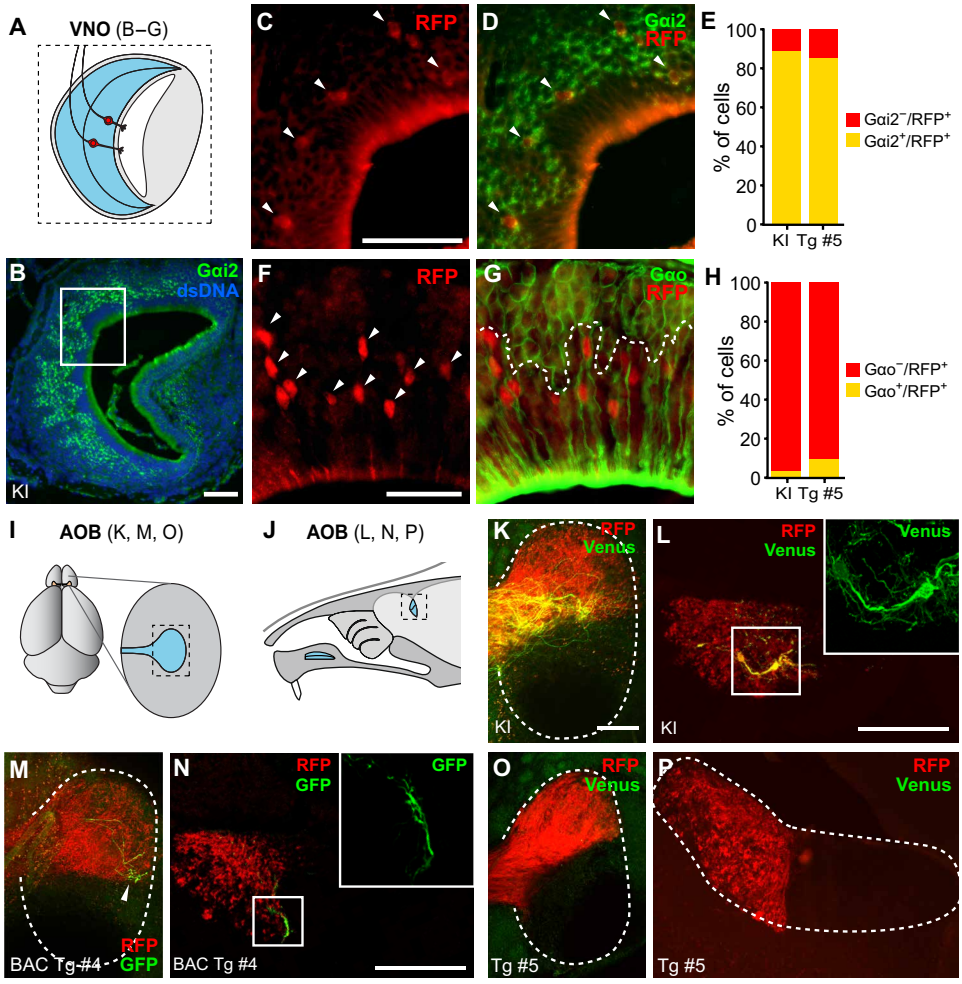

Fig. 6. Transgenic Fpr-rs3 expression is punctate, restricted to Gαi2-expressing neurons, and exclusively present in neurons targeting the rostral AOB. (**A**) Schematic of a representative VNO coronal section with the neuroepithelium highlighted in light blue and two apical VSNs in red. (**B**) In situ hybridization with a probe against Gαi2 (green) and counterstained with DAPI (blue), on a VNO coronal section of a *Fpr-rs3*$^{cre/+}$*; Rosa*$^{stopRFP/+}$ mouse. The white square shows the magnification displayed in (C) and (D). Scale bars, 100 μm. (**C**) Magnification of (B) with immunolabeling of RFP (red). White arrowheads point at RFP$^+$ VSNs. (**D**) Same image as (C) showing Gαi2 RNA localization in overlay. (**E**) Numbers of cells that are either negative for Gαi2 and positive for RFP (Gαi2$^-$/RFP$^+$; red) or double positive (Gαi2$^+$/RFP$^+$; yellow) in VNO sections of *Fpr-rs3cre/+; Rosa*$^{stopRFP/+}$ *mice* ($n = 10$ sections) or *Tg(Fpr-rs3-cre)5; Rosa*$^{stopRFP/+}$ (Tg #5, $n = 10$ sections) mice. (**F** and **G**) In situ hybridization with a probe against Gαo (green) and immunolabeling of RFP (red) on a VNO coronal section of a *Fpr-rs3cre/+; RosastopRFP/+* mouse. Dashed line delimits the region where cell bodies are Gαo$^+$. White arrowheads point at RFP$^+$ VSNs. Scale bar, 50 μm. (**H**) Numbers of cells that are either negative for Gαo and positive for RFP (Gαo$^-$/RFP$^+$; red) or double positive (Gαo$^+$/RFP$^+$; yellow) in VNO sections of KI ($n = 15$ sections) or Tg #5 (Gαo, $n = 26$ sections) mice. (**I** and **J**) Schematic of the AOB on a dorsal view of a mouse brain (I) and location of the AOB on a lateral view of a mouse head hemisection (J). Dashed boxes indicate areas displayed in (K) to (P). (**K** to **P**) Endogenous protein fluorescence (RFP, red; Venus/GFP, green) in a *Fpr-rs3cre/+; Rosa*$^{stopRFP/+}$ mouse (**K** and **L**), a *Tg(BAC-Fpr-rs3-cre)4; Rosa*$^{stopRFP/+}$ mouse (**M** and **N**), and (**O** and **P**) a *Tg(BAC-Fpr-rs3-cre)5; Rosa*$^{stopRFP/+}$ mouse. Scale bars, 200 μm. In addition, see fig. S1 (A to D).

clusters, despite very high rates of gene duplication of V1r/Fpr genes, whose evolutionary dynamics are closer to those of LINE elements than other clustered genes such as Hox or opsin genes. This clustered organization contrasts with the one of pseudogenes, which are often found as singletons. We also observed that V1r transcriptional levels are different among V1r gene clusters and that the presence of pseudogenes inside V1r gene clusters is positively associated with their transcriptional activity. These data suggest the existence of regulatory sequences in V1r/Fpr gene clusters that are necessary for adequate vomeronasal receptor expression. These regulatory sequences, acting at distance and possibly shared by multiple vomeronasal genes, would be lost when a clustered V1r gene duplicates and lands outside a V1r gene cluster, eventually resulting in the pseudogenization of the now singleton V1r. The rare exceptions to this model would reflect duplication events that included the minimal regulatory sequences necessary for functional V1r/Fpr transcription. Our hypothesis—the inability of nonlocal single gene duplications to recapitulate normal transcription—thus explains the evolutionary conserved clustered organization of functional vomeronasal genes.

We tested our hypothesis by performing a gene transplantation experiment. Our results strongly suggest that a V1r/Fpr transgene that is located outside of its native V1r/Fpr cluster can be chosen for expression but cannot maintain transcriptional activity. An alternative explanation of our data would be that the transgenes are subject to nonspecific (leaky) expression. Multiple observations argue against this explanation. First, we observed temporally specific

expression. Transcription of the transgene is limited to young neurons, the time period at which chemoreceptor gene choice takes place. Second, we observed cell type–specific expression. The transgenes are selectively expressed in apical (Gαi2-expressing) VSNs that project to the rostral part of the accessory olfactory bulb. Third, we observed high levels of transgene expression that are as high as endogenously expressed vomeronasal genes, something that one would not expect from leaky expression. Last, we did not observe neurons dying after initiating transcription of the transgene. Our data thus point to a transgene activation that is punctate, strong, transient, and specific to a developmental phase and to a particular identity of vomeronasal neurons. They support two novel and unexpected findings: (i) that V1r/Fpr genes are bound to their clusters because their adequate expression depends on this proximity and (ii) that the clustered organization supports transcriptional stabilization rather than initial gene choice.

Why is it that the critical role played by vomeronasal gene clustering on transcriptional stability was not reported long ago? A first answer surely lies in the fact that most transgenic-based approaches that do not recapitulate endogenous expression are usually never published. This is in fact a situation that we encountered years ago when trying to drive the expression of a fluorophore under the control of a vomeronasal receptor promoter (*V1rb2*). We observed very rare transgene expressing neurons in the VNO, expression that disappeared with age. This finding, ignored at the time because not following our expectations, parallels the experiments reported here. However, there is one report that is based on a transgenic approach to drive expression of fluorophores under the control of vomeronasal receptor genes (*12*). In this work, the observed expression pattern recapitulated the endogenous pattern of the corresponding receptor genes. An easy explanation of this apparent unexpected ability lies in the size and the contents of the transgene used in the study: a very large BAC containing six complete V1r genes; in other terms, a small V1r gene cluster.

Our data are reminiscent of and congruent with previous work in which we identified a peculiar mechanism affecting V1r genes that we termed the "gene cluster lock" (*13*). When a nonfunctional V1r allele is chosen for expression, the VSN goes on to coexpress another random V1r allele (a situation likely similar to the receptor switch following transgene silencing we encountered in the present study). This second V1r choice appears to include any V1r gene, but not genes from the original chosen cluster allele. This "cluster lock" is naturally suggestive of regulatory elements that may act on multiple receptor genes, but one gene at a time, in a given cluster. The nature of these regulatory elements is still to be defined and could, for example, involve a limited number of sequences present in vomeronasal gene clusters, acting in cis on vomeronasal genes. The unusually high homology between V1r promoters pertaining to a given subfamily could reflect these latter being a target for these modulatory elements.

Similar to V1rs and Fprs, Ors expressed in the main olfactory epithelium are also subject to random monogenic gene choice (*14*). Various reports have shown the role played by cis-regulatory elements present in olfactory gene clusters, elements that are critical for olfactory gene expression. These cis-regulatory sequences are currently considered as choice elements. It is naturally tempting to draw mechanistic parallels between these data and our current findings, but one should remain cautious because large differences exist between both odorant and vomeronasal receptors. First, Ors and Vrs/Fprs share no sequence homologies, both at the levels of their corresponding genes and proteins. Possibly more important relative to the regulation of their expression, V1r and Fpr genes exhibit highly conserved promoters (*15*, *16*), at least within each subfamily, a feature completely lacking in Or genes. Last, as demonstrated in this work, Vrs are largely more diverse than Ors between species, as a result of a high rate of gene birth and death.

Notwithstanding the large number of vomeronasal receptor genes in mammalian genomes and despite the critical role played by their products in species survival, vomeronasal receptor gene regulation is not understood. We provide here, in addition to the identification of a selective pressure that maintained vomeronasal genes clustered during evolution, with the idea that the elements lost during nonlocal, single vomeronasal gene duplications regulate transcritption and more precisely the stabilization of gene choice rather than the frequency of the choice itself. This latter idea represents a novel direction of research in the field.

MATERIALS AND METHODS

Genomic mapping of V1r genes

De novo CDS retrieval

For each species, amino acid sequences of the given repertoire were retrieved by using BioMart on the Ensembl (*17*) website and by specifying the corresponding InterPro family in the filter field "Limit to genes with these family or domain IDs." The following InterPro signatures (*18*) were used: IPR000725 (Ors), IPR004072 (vomeronasal receptor type 1), and IPR000826 (Fpr receptor related). Then, this set of sequences was used as a TBLASTN query to identify de novo the repertoire CDSs. To avoid redundant searches with the TBLASTN method, the query set was reduced to sequences sharing no more than 80% of identity using BLASTP. TBLASTN searches were performed on genome assemblies listed in Table 1 using the best-hit filtering algorithm (options -best_hit_score_edge 0.3 and -best_hit_overhang 0.3) and keeping only hits with an E value below 1×10^{-20}. Overlapping hits were merged, and putative CDSs were retrieved. The CDSs were translated and aligned using the MAFFT program (*19*) version 7.4 with the G-INS-i algorithm for the curation process. The curation consisted in manual review of the multiple alignment of protein sequences from each repertoire. During that process, we eliminated sequences that had too many ambiguous positions (>30 N), that had a stop codon before the last conserved motif, that had gaps or mutations in any of the conserved motives, or that had a large (>30 bp) deletion.

Pseudogene sequence retrieval

The validated sets of olfactory receptor CDSs were used to retrieve pseudogenes from the genome assemblies using a reciprocal BLAST. For this, a first round of BLAST was performed with the olfactory receptor–curated CDSs of a given species onto its genome assembly using the discontigous megablast algorithm. From this search, self-hits were excluded, and only hits below an E value of 1×10^{-4} were retained. In addition, query sequences that hit more than 50 genomic locations were considered as repeated elements and discarded. The sequences found with this first round were BLASTed onto the CDS collection of the species of interest, and only the sequences whose best hit was a gene from the targeted olfactory receptor family were considered as homologous pseudogenes. In addition, the pseudogene sequences that were closer than 6500 bp were joined and aligned to their closest intact receptor CDS. If an intact CDS could cover most of the joined sequences, then we considered that

Table 1. Species names, corresponding genome assemblies, and number of sequences retrieved for each repertoire. The bold names correspond to the species analyzed in Fig. 1 (C to F). For the sequences that were identified by polymerase chain reaction from biopsies (cf. part II), the genome assembly field is not applicable (n.a.). Asterisks in parentheses show species for which more than one assembly was used. Under each repertoire column (Or, odorant receptor repertoire; V1r, type 1 vomeronasal receptor repertoire; Fpr, formyl peptide receptor repertoire), the number of intact CDSs is indicated. When available, the number of pseudogenes (ps) is added after a slash (/). Blank cells correspond to repertoires that were not identified.

Species	Assembly	Or	V1r (/ ps)	Fpr (/ ps)
Ailuropoda melanoleuca	ailMel1	641	11	
Bos taurus	ARS-UCD1.2	1066	36 / 46	
Callithrix jacchus	ASM275486v1	350	7	
Canis familiaris	CanFam3.1	814	9	
Cavia porcellus	cavPor3	805	91	
Choloepus hoffmanni	choHof1	496	10	
Dasypus novemcinctus	Dasnov3.0	1513	51	
Dipodomys ordii	Dord_2.0	497	36	
Equus caballus	EquCab3	1040	37 / 71	3 / 2
Erinaceus europaeus	eriEur1	534	31	
Felis catus	Felis_catus_6.2	680	21	
Ictidomys tridecemlineatus	SpeTri2.0	929	68	
Loxodonta africana	Loxafr3.0	1925	34 / 94	1 / 0
Microcebus murinus	Mmur_1.74	356	72	
M. domestica	monDom5(*)	1188	93 / 18	4 / 2
Mus musculus	GRCm38/mm10	1140	235 / 159	7 / 2
Mustela putorius furo	MusPutFur1.0	817	11	
Ochotona princeps	pika	344	50	
Ornithorhynchus anatinus	OANA5	262	302	
Oryctolagus cuniculus	OryCun2.0	772	154	
Otolemur garnettii	OtoGar3	751	43	
Ovis aries	Oar_v3.1	459	27 / 43	0 / 0
Procavia capensis	proCap1	391	16	
R. norvegicus	RGSC3.4(**)	1350	104 / 118	6 / 0
Sarcophilus harrisii	Devil_ref v7.0	945	73	
Sorex araneus	COMMON_SHREW1	639	39	
Sus scrofa	Sscrofa10.2	1278	13	
Tupaia belangeri	tupBel1	555	37	
Vicugna pacos	vicPac1	348	13	

(*)The Or repertoire of the *Monodelphis domestica* was identified in BROAD5. (**)The Or repertoire of *Rattus norvegicus* was identified in Rnor_6.0.

the parts of the joined sequences belonged to the same pseudogene that was split by insertions.

Gene cluster and singleton definition

To assign V1R and FPR genes and pseudogenes to clusters, we took the following approach, that was applied independently for each species. First, each gene or pseudogene was attributed a single coordinate, which is the 5′ genome coordinate of either the intact CDS for genes or the pseudogenic CDS for pseudogenes. Next, we calculated all the distances between adjacent genes. The sorted distances were then split into two groups using the Jenks natural break optimization (Jenks, 1967) for $k = 3$ (*20*). In that manner, the middle break was used to unbiasedly separate two categories of distances: the smaller distances representing the intracluster distances and the longer distances representing the intercluster distances. Next, we calculated the mean and the SD of the intracluster distances and defined the clustering threshold as the mean plus three times the SD. Last, clusters were defined by aggregating neighboring genes whose attributed coordinates were closer to each other than the clustering threshold. Genes that could not be assigned to any cluster because they were too far from another gene or because they were alone on a chromosome or a scaffold were considered as singletons.

Sequence alignment and phylogeny

Twenty-nine mammals species phylogeny

The mammalian phylogeny presented in Fig. 1A was based on the phylogeny published by Meredith *et al.* (*21*).

Twenty-nine mammalian OR and V1R phylogenies

For the 22,885 OR genes, we proceeded as followed. First, a multiple alignment of OR protein sequences was obtained independently for each species with MAFFT using the G-INS-i algorithm. In these alignments, sites with more than 99% of sequences presenting a gap were removed. The sequences were then dealigned, pooled in a single file, and realigned with MAFFT using the options --retree 2 and --maxiterate 0. The resulting alignment was manually trimmed to keep the sequences in between the first conserved asparagine and the first conserved motif appearing after the last transmembrane domain. Again, sites with more than 99% of sequences presenting a gap were removed. To root the phylogeny, we aligned five adrenergic receptor beta 2 orthologs to serve as outgroup and merged this alignment with the OR alignment using the --seed option and the G-INS-i algorithm.

For the 1724 V1Rs, the overall alignment procedure was the same except (i) V1R sequences from all species were directly pooled and aligned with the options --retree 2 and --maxiterate 0 and (ii) the motives that we used to trim the alignment. In this step, we kept the sequences between the first conserved methionine, which correspond to the start methionine of most de novo–retrieved V1R CDS, and the end of the last transmembrane domain, whose sequence identity is conserved among V1Rs. To root the phylogeny, we used five type 2 taste receptor orthologs that we merged to the V1R alignment using the same method as for the ORs.

The resulting protein alignments were used for phylogenetic reconstruction with FastTree v2.1.10 (*22*), specifying the options -gamma, -spr 4, -mlacc 2, and -slownni. The trees were uploaded on the Interactive Tree Of Life webtool (https://itol.embl.de/) and rerooted manually using the aforementioned outgroups.

Inparalog group size quantification

To count the number of genes within each inparalog group of the OR and the V1R phylogenies (Fig. 1, A and B), we used a recursive algorithm that split the rooted gene phylogeny until producing subtrees with gene from a single species. The number of genes in these subtrees was sorted by category of size for representation in Fig. 1B.

RNA sequencing

Sample preparation and sequencing for bulk tissue RNA-seq

Eight-week-old C57BL/6J mice ($n = 4$ males and $n = 4$ females) and 8-week-old Lewis male rats ($n = 7$) were used for bulk RNA-seq. Vomeronasal epithelia were isolated; transferred to a tube containing 500 to 600 µl of ice-cold lysis buffer, 5 to 6 µl of β-mercaptoethanoland, and a 0.5-cm-diameter steel ball; and then put on ice following the manufacturer's protocol (QIAGEN). Samples were then homogenized using a FastPrep-24 instrument (MP Biomedicals) at 6 m/s for 30 s and kept on ice. RNA was extracted using the QIAGEN RNeasy Mini Kit SII following the manufacturers' instructions. Two deoxyribonuclease (DNase) treatments were performed on all samples: first using the QIAGEN RNase-Free DNase Set and then using the Life Technologies Ambion DNase I Kit. RNA samples were then aliquoted and stored at −80°C.

Complementary DNA (cDNA) libraries were generated with the TruSeq RNA and DNA sample preparation kits after selection of poly A–containing mRNAs. Adapters for RNA-seq multiplexing were added to the cDNAs. The cDNA libraries were sequenced with the HiSeq 2500 Sequencing System to generate stranded single-end 100 reads of 100 nucleotides in length.

Sample preparation and sequencing for scRNA-seq

For scRNA-seq, $OMP^{GFP/wt}$ males aged of 67 days were used ($n = 7$). VNOs were dissected, and their cartilaginous envelope was discarded. The tissues were immediately placed in ice-cold oxygenated artificial cerebrospinal fluid containing the following: 124 mM NaCl, 3 mM KCl, 2 mM $CaCl_2$, 1.3 mM $MgSO_4$, 26 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, and 10 mM D-glucose with an osmolarity of 300 mosmol and pH 7.4 when oxygenated with 95% O_2 and 5% CO_2.

The method used for the preparation of samples for scRNA-seq was described in details previously (*23*). Following microdissection, the tissues extracted from the seven mice were pooled into two tubes (i.e., three and four mice, respectively). Tissue dissociations were performed using the Papain Dissociation System (catalog no. LK003150, lot #35S16330; Worthington Biochemical Corporation, NJ, USA) following the manufacturer's protocol, with slight modifications; the EBSS solutions and the medium solution with serum were adapted from Saxena *et al.* (*24*) The duration of tissue dissociation was of 20 min at 37°C.

To ensure the exclusive isolation of live nucleated cells, cell suspensions were incubated with Hoechst 33342 [2 µg/ml; an ultraviolet (UV) fluorescent adenine-thymine binding dye; #H1399, Life Technologies] at 37°C for 15 min. To exclude dead cells, 1 µM DRAQ7 (a far-red fluorescent DNA intercalating dye; #DR71000, BioStatus) was added to the cell suspensions before fluorescence-activated cell sorting (FACS). GFP^+/$Hoechst^+$/$DRAQ7^-$ cells were then sorted in an empty Eppendorf tube according to their forward scatter (FSC) and side scatter (SSC) properties using a Beckman Coulter MoFlo Astrios (Miami, FL) cell sorter with a 100-µm nozzle at a pressure of 25 psi. Doublets were excluded after gating on FSC-A/FSC-H, followed by SSC-H/SSC-W. Approximately 4000 GFP^+/$Hoechst^+$/$DRAQ7^-$ cells were collected from each single-cell suspension pool, each in a final volume of 10 µl.

After FACS sorting, 4 µl of C1 Suspension Reagent (Fluidigm) was added to the 10 µl of cell suspensions, yielding mixes of approximately 300 cells/µl. Cells were captured using the C1 Single-Cell mRNA-seq high-throughput (HT) integrated fluidic circuit (IFC) designed for 10 to 17 µm of cells (catalog no. 100-5760, Fluidigm). Each 14 µl of mix was loaded on one side of the chip that was processed on the C1 System (catalog no. 100-7000, Fluidigm) following the manufacturer's protocol. Following cell capture, the chip was imaged using an automated inverted fluorescent microscope (Axio Observer Z1, Zeiss) that allowed to locate single GFP^+ cells in the 800 capture chambers. A custom MATLAB script was used to interpolate and calculate the position and the focus in z for each of the 800 capture chambers based on those defined for the first capture chamber and about 20 capture chambers randomly distributed across the C1 chip. Cell lysis, barcoding, reverse transcription, and polymerase chain reaction amplification were performed directly on the C1 chip using the Fluidigm's C1 Single-Cell mRNA Seq HT Reagent Kit v2 (catalog no. 101-3473, Fluidigm) following the manufacturer's protocol. The cDNA content of each column on the chip (formed of 40 cells) was pooled before harvesting. Twenty cDNA libraries were generated using the Nextera XT DNA Library Preparation Kit (catalog no. FC-131-1024, Illumina) and the Nextera XT Index Kit (catalog no. FC-131-1002, Illumina) following Fluidigm's protocol. Before the final purification step, all 20 cDNA libraries were pooled. The library pool molarity and quality were assessed with the Qubit 2.0 using the Qubit dsDNA HS Assay Kit (catalog no. Q32854; Thermo Fisher Scientific) and the TapeStation using the

Agilent High Sensitivity DNA chip (catalog no. 5067-5584, Agilent Technologies). The pool was then loaded at 7 pM on four lanes for clustering on a rapid paired-end Illumina flow cell (catalog no. PE-402-4002, Illumina) and sequenced on an Illumina HiSeq 2500 sequencer using the HiSeq Rapid SBS Kit v2 (catalog no. FC-402-4021, Illumina) chemistry. Read 1 consisted of 11 bases (six bases for the cell barcode and five bases for the unique molecular identifier), while read 2 was formed of 80 bases.

Custom gene annotation and sequence read mapping

To increase the sensitivity of RNA-seq read assignment to vomeronasal receptor genes (V1rs, V2rs, and Fprs), we manually annotated their transcript 3′ untranslated regions (3′UTRs), which were mostly absent or incomplete in the available mouse genome annotations. To this aim, we retrieved the Ensembl transcript annotations for each gene to be updated. We selected a single Ensembl transcript given the following criteria: (i) does not form splice isoform with other genes; (ii) has a CDS annotation; and if not, (iii) encompasses the gene CDS. With the Integrative Genomics Viewer software (*25*), we visualized the read alignment coverage of each gene's last exon to determine the 3′ end of the transcript. For each gene whose 3′UTR could not be clearly identified because of a lack of read coverage, the closest homologs were collected with BLAST, and the 3′UTR length was deduced from the mean of the homologous gene 3′UTR lengths. For each updated gene, we replaced all previous transcript isoforms with the novel transcript annotation in the Ensembl version 86 GTF file of the mouse genome GRCm38 (mm10). De novo–identified V1r pseudogenes were added as well to our custom gene annotation. Exceptions are the pseudogenes *Vmn1rn-ps47*, *Vmn1rn-ps50*, and *Vmn1rn-ps51* that are nested in functional V1r 3′UTRs and therefore not quantifiable as individual transcripts. High-sequence redundancy leads to ambiguous read mapping, which can hamper RNA-seq quantification if reads assigned to multiple loci (multimapping reads) are discarded. On the other hand, allowing multimapping reads leads to artificial expression level estimation. The V1r family D (V1rD), which is clustered (*6*), has recently expanded in the *Mus* lineage, leading to high-sequence redundancy between genes. To optimize the maximum allowed number of multimapping reads to estimate the expression levels of V1rD genes, we compared all 8-kb nucleotide sequences starting at each start of the CDSs found in this cluster and extended in the corresponding orientation. Figure S2A shows the organization of V1rD gene sequences grouped by homology. To estimate how sequence redundancy within the V1rD family may affect RNA-seq read assignment, we calculated the copy number of every 100 nucleotides sequence found in the 8-kb sequence set representing V1rD transcripts. With this approach, we estimated that only 57% of RNA-seq reads from V1rD transcripts could be assigned without ambiguity (fig. S2B). Therefore, to assess the expression of V1rD genes, we chose to retain a maximum of four ambiguous assignments for each read for which we estimated to retain 91.6% of read alignments (fig. S2B). RNA-seq reads were mapped with STAR (*26*) version 2.7.1a, setting the parameters --outFilterMultimapNmax at 4.

Transcript quantification

Gene expression quantification of the bulk RNA-seq data was carried using featureCounts (*27*) version 1.6.5 and our custom gene annotation file, considering reversely stranded reads (-s 2); fractional counting of multimapped reads was performed (-M --fraction). For the scRNA-seq data, digital gene expression matrices were generated using UMI–tools version 1.0.0 (*28*) following the steps described previously (*23*).

Data filtering (scRNA-seq)

Single-cell RNA-seq analyses were performed on R version 3.5.0. Capture chambers containing a single GFP^+ cell were selected after visual inspection. The transcriptomes contained in capture chambers characterized as empty were checked and added to the final set of single-cell transcriptomes when they displayed high expression of an olfactory chemoreceptor gene, an expression comparable to the ones of capture chambers containing cells. These purportedly empty capture chambers contained a similar number of genes, abundance of mitochondrial gene counts and sequencing depth than those of the selected chambers. This resulted in the collection of 412 single-cell transcriptomes (394 + 18 retrieved from visually empty chambers). Additional cell filtering criteria were applied following Mayer *et al.* (*29*), as described previously (*23*). (i) We removed all cells characterized by less than 1000 expressed genes. (ii) We removed all cells for which mitochondrial counts exceeded 10% of their total counts, as high mitochondrial counts indicate suffering or dead cells. (iii) We removed all cells for which the total number of reads, the total number of detected genes, the total number of unique molecular identifiers (UMIs), and the percentage of mitochondrial counts were three median absolute deviations away from the median, after $\log_{10}$ transformation. (iv) We removed all cells that showed unusually high or low number of UMIs given their number of reads, after $\log_{10}$ transformation, by fitting a LOESS curve (loess function of the stats R package, with a span = 0.5 and a degree = 2) where the number of UMIs is taken as response variable and the number of reads as predictor. Those cells for which the model residual was not within three median absolute deviations of the median were filtered out. (v) We also removed all cells that showed unusually high or low number of genes given their total number of UMIs, also after $\log_{10}$ transformation, following the above-mentioned criteria. No cells were removed after step 1, and only one cell was removed after step 2. However, about 11.2% of the cells were removed during steps 3 to 5, which left 365 cells for downstream analyses. No gene filtering was applied on this dataset.

VSN identification (scRNA-seq)

For the estimation of expression levels, raw gene expression values (UMI counts) were normalized by the total number of counts per cell and scaled to 10^4. Because of technical limitations associated to the Fluidigm C1 HT IFC chip, VSN transcriptomes did not display monogenic and monoallelic expressions of olfactory receptor genes. We speculate that this technical artifact results from the fact that these chips suffer from leakages of mRNA molecules between capture chambers, which is exacerbated by the strong expression of olfactory receptor genes in VSNs. Nevertheless, we attributed to each VSN the identity of the olfactory receptor gene that displayed the highest expression. With that, we identified 267 cells with a functional V1r or apical Fpr identity.

Knockin and transgenic lines

Fpr-rs3$^{iVenus\text{-}iCre}$, KI

This allele consists in a targeted modification of the 3′UTR of *Fpr-rs3*, in which the sequence IRES-tauVenus-IRES-Cre was inserted to produce a tri-cistronic transcript.

Tg(Fpr-rs3-iVenus-iCre), Tg

This allele consists of a minimal *Fpr-rs3* gene bearing the same modification as the knockin allele. To generate this transgene, we built a

vector containing *Fpr-rs3* promoter (400 bp), first exon (53 bp), the 5′ end (1163 bp) and the 3′ end (784 bp) of the original 24-Kb intron, the 3′ end of the original intron, and, from here, the same sequence as the knockin allele (*Fpr-rs3* CDS, IRES-tauVenus-IRES-Cre sequence, and 1977 bp of the *Fpr-rs3* 3′UTR). The intron of this short transgene was reduced (chr17: 20624888-20626051 joined to chr17: 20648228-20649012 in GRCm38) due to plasmid size limitations. This vector was injected in the pronuclei of C57BL6/J-DBA/2J F2 zygotes. Nine founders were obtained, five lines were characterized, and three lines were analyzed in this work, labeled as #2, #5, and #7.

Tg(Fpr-rs3-itCre-iGFP), BAC Tg

This allele consists of a *Fpr-rs3-IRES-Cre-IRES-GFP* containing BAC transgene. To generate this transgene, we modified the BAC RP23-80N19, comprising the entire *Fpr-rs3* gene, in which we inserted the cassette IRES-Cre-IRES-GFP-frt-Zeo-Puro-Frt after *Fpr-rs3* stop codon. The resistance cassette was removed by in vitro flipase-induced recombination, and the resulting vector was injected in the pronuclei of C57BL6/J-DBA/2J F2 zygotes. Two independent lines were analyzed in this work, labeled as #2 and #4.

Rosa^{stopRFP}

KI, Tg, and BAC Tg mice were crossed with *Rosa26^{RFP}* mice (*30*) on a C57BL/6J background.

Histological preparations

Tissue fixation

Olfactory bulbs were fixed in vivo through cardiac perfusion of 10% formalin (Sigma-Aldrich) before dissection. For this, mice were deeply anesthetized with pentobarbitalum natricum at 150 mg/kg (Streuli Pharma) and perfused transcardially with phosphate-buffered saline (PBS)–heparine (20,000 UI/liter; Bichsel) (pH 7.4). After flushing, mice were perfused with 10% formalin (Sigma-Aldrich). Dissected organs were immersed in the same fixative and stored overnight at 4°C. VNOs used for immunohistochemistry, standard in situ hybridization, and quantification of endogenous fluorescence were directly dissected and immersed in fixative overnight at 4°C. For RNAscope assays, VNOs were immersed in fresh 4% paraformaldehyde (PFA) overnight at 4°C.

Cryosections

Except for olfactory bulbs, all histological analyses were performed on cryosections. VNOs used for immunohistochemistry and standard in situ hybridization were transferred into a 15% sucrose PBS solution for 12 hours at 4°C and then into a 30% sucrose PBS for 12 hours at 4°C. Sucrose-treated tissues were embedded in optimal cutting temperature (OCT) compound and frozen. Cryosections (14 to 20 μm) were cut with a cryostat and mounted on Superfrost Plus slides (Thermo Fisher Scientific). Slides were stored at −80°C. For VNO sections used in RNAscope assays, the procedure was the same except that sections were consecutively immersed in 10% sucrose, 20% sucrose, and 30% sucrose, each for 12 hours at 4°C. Sections were cut at 16 μm.

Vibratome sections

For olfactory bulbs, fixed organs were embedded in agarose low melt (Thermo Fisher Scientific), and sections (50 μm) were cut with a vibratome. Sections were postfixed for 10 min in 10% formalin at room temperature, then washed, and immersed in PBS. Sections were either stored at 4°C for a maximum of 1 week or directly processed.

Stainings

Standard in situ hybridization

RNA probes against vomeronasal Fpr transcripts (*Fpr-rs3*, *Fpr-rs4*, *Fpr-rs6*, and *Fpr-rs7*); V1r families A, E, and J (*V1ra3*, *V1re4*, and *V1rj3*); *Vmn2r1*; and *Gαi2* are described in previous publications (*6*, *13*) (see Table 2).

Digoxigenin (DIG)–labeled RNA probes were prepared using a DIG RNA labeling kit (Roche) following the manufacturer's instructions. Unless specified, all steps are performed at room temperature. Vomeronasal cryosections (14 μm) were dried for 40 min, postfixed in 10% formalin for 15 min, and washed twice with 1× PBS for 3 min. Slides were incubated in 0.1% H_2O_2 for 30 min and washed twice with 1× PBS for 3 min. Slides were incubated with proteinase K (10 μg/ml) diluted in TE [10 mM tris (pH 7.5) and 1 mM EDTA (pH 8)] for 5 min, fixed another time with 10% formalin for 10 min, and washed twice with 1× PBS for 3 min. Slides were immersed in 0.2 M HCl for 10 min, washed with 1× PBS for 3 min, and incubated in 0.1 M trethanolamine-HCl (TEA-HCl) (pH 8) for 1 min. Slides were transferred into a fresh solution of TEA-HCl with 0.25 μl of acetic anhydre per milliliter of TEA for 10 min and washed twice with 1× PBS for 2 min. Hybridization buffer was prepared on ice with 1× salt buffer (10× stock solution: 2 m NaCl, 100 mM tris, 50 mM $NaH_2PO4{\cdot}H_2O$, 50 mM $Na_2HPO_4{\cdot}2H_2O$, 0.5 M EDTA, HCl to adjust to pH 7.5, and H_2O-DEPC), 50% formamide, 10% dextran sulfate, tRNA (1 μg/μl), and 1× Denhardt's. Probes were denatured for 7 min at 70°C and diluted into the hybridization buffer at 500 ng/ml. Hybridization buffer was added onto slides, which were then covered with a sterile coverslip, put in a humidified, sealed box, and kept at 62°C for 14 hours of hybridization. Following hybridization, slides were washed twice at 62°C and once at room temperature with preheated washing solution (1× saline sodium citrate, 50% formamide, 0.1% Tween 20, and H_2O). Slides were preincubated in 1× maleic acid buffer containing tween 20 (MABT) (5× MABT: 500 mM maleic acid, 750 mM NaCl, 0.5% Tween 20, and H_2O) and 2% Roche-blocking reagent (10% blocking reagent and 1× MABT) for 30 min.

Anti–DIG-alkaline phosphatase

Slides were incubated 1 hour with anti–DIG–alkaline phosphatase–conjugated antibody (Roche, 11093274910) diluted 1:1000 in preincubation mix. Slides were washed 3× 5 min with tris NaCl

Table 2. ISH probes. nt, nucleotide.

Target	Position of nucleotides relative to the start ATG	Origin	Labeling
Gαi2	nt 248–1138	Mouse	DIG, FastRed
Vmn2r1	nt 1898–2733	Mouse	DIG-peroxydase, streptavidin
Fpr-rs3	nt 710–1057	Mouse	DIG, FastRed
Fpr-rs4	nt 841–1080	Mouse	DIG, FastRed
Fpr-rs6	nt 854–1226	Mouse	DIG, FastRed
Fpr-rs7	nt 967–1268	Mouse	DIG, FastRed
V1ra3	nt 249–839	Mouse	DIG, FastRed
V1re4	nt 178–754	Mouse	DIG, FastRed
V1rj3	nt 334–775	Mouse	DIG, FastRed

Tween (TNT) (150 mM NaCl, 100 mM tris, 0.05% Tween 20, and HCl to pH 7.5). Alkaline phosphatase activity was detected by incubating slides with FastRed substrate (Dako, K0597) for 30 min. Slides were rinsed with running water for 2 min.

Anti–DIG-peroxidase

Slides were incubated 1 hour with anti–DIG-peroxidase–conjugated antibody (Roche, 11207733910) diluted 1:100 in preincubation mix. Slides were washed three times with TNT [150 mM tris, 150 mM NaCl, and 0.05% Tween 20 (pH 7.5)], incubated for 30 min with a biotinyl-tyramide solution (PerkinElmer), again washed three times with TNT, and incubated for 30 min with streptavidin–*Alexa Fluor 488* (Molecular Probes). Slides were rinsed with running water for 2 min. Slides are kept in 1× PBS and immediately proceeded to the preincubation step in the immunohistochemistry protocol (see below).

Immunohistochemistry

Cryosections stored at −80°C were thawed and dried 30 min at room temperature, then rinsed with PBS and 0.5% Triton X-100 for 5 min, and placed in a humid immunohistochemistry (IHC) box. Sections were preincubated with PBS, 0.5% Triton X-100, and 10% fetal calf serum for 1 hour at room temperature. Sections were incubated with the primary antibody (see Table 3) diluted in PBS, 0.5% Triton X-100, and 10% fetal calf serum at 4°C for 12 hours. After incubation, sections were washed 3× for 15 min in 1× PBS and 0.5% Triton X-100 and then incubated with the secondary antibody (see Table 4) diluted in PBS, 0.5% Triton X-100, and 10% fetal calf serum at room temperature for 90 min. Sections were again washed 3× for 15 min in 1× PBS and 0.5% Triton X-100. Sections were counterstained with 4′,6-diamidino-2-phenylindole (DAPI; 1:5000) for 5 min, rinsed with 1× PBS, and mounted with 1,4-diazabiclo(2.2.2) octane (DABCO) (Sigma-Aldrich) in glycerol.

Evaluation of the rate of reporter-expressing cells across apical and basal VSN populations

To estimate the proportion of different chemoreceptor populations among RFP^+ neurons, different in situ hybridization (ISH) probes specific to either V1r or V2rs transcripts were used to cover a broad range of VSN identities from the apical and the basal layer of the sensory epithelium (see Table 1, ISH probes). For each ISH chemoreceptor probe, the percentage of labeled RFP^+ neurons was calculated. To normalize this value relative to the probe coverage, the coexpression percentage was divided by the mean number of labeled neurons per VNO section (a population size factor). The resulting values were summed by group of probes (namely, V1rs, V2rs, and Fprs) and displayed as relative proportions for each mouse line, representing the proportion of apical and basal VSN neurons having expressed the transgenic *Fpr-rs3* alleles.

Cell death rate quantification

Caspase 3 fluorescent IHC staining on VNO cryosections was imaged with a Leica DM5500 epifluorescence microscope, with the 63× objective, and images centered on single transgene-expressing cells were acquired. For quantification, we defined a 50-μm radius around the transgene-expressing cell and counted all cells within this area based on the DAPI staining. The cell death rate was estimated from the number of Caspase 3–labeled cells over the total number of either transgene-expressing cells or neighboring cells.

RNAscope single-molecule in situ hybridization

Fifteen- to 17-day-old male and female *Fpr-rs3*$^{iVenus\text{-}iCre}$, Tg(Fpr-rs3-itCre-iGFP) line #2 and line #4 mice were euthanized, and VNOs were removed and fixed in freshly prepared 4% PFA overnight at 4°C. Tissue was sequentially immersed in 10% sucrose for 12 hours, 20% sucrose for 12 hours, and 30% sucrose for 12 hours, always at 4°C. Heads were embedded on OCT, frozen in liquid nitrogen, and stored at −80°C until sectioning. The VNO was cut on a cryostat microtome in 16-μm coronal sections. For each mouse, about 25 sections were used for RNAscope ISH. RNAscope staining was performed according to the manufacturer's protocol (RNAscope Multiplex Fluorescent V2 Assay, reference no. 323136, Advanced Cell Diagnostics). Pretreatment was performed according to guidelines for fixed frozen tissue and included postfixation, dehydration, hydrogen peroxide treatment, 5-min target retrieval, and 3-min protease III

Table 3. Primary antibodies. IgG, immunoglobulin G.

Antibody	Provider	Concentration	Secondary antibody
Rabbit polyclonal IgG anti–FPR-rs3	Eurogentec (custom design)	1:1500	#1
Rabbit polyclonal IgG anti-RFP	Abcam (reference no. ab34771)	1:500	#2
Rabbit polyclonal IgG anti-Gnao1	GeneTex (reference no. GTX114439)	1:100	#3
Chicken polyclonal IgG anti-RFP	Rockland (reference no. 600-901-379)	1:500	#4
Rabbit monoclonal IgG anti–Caspase 3	BD Biosciences (reference no. 559565)	1:500	#1

Table 4. Secondary antibodies.

Antibody	Conjugate	Provider	Concentration	Number
Goat polyclonal IgG anti-rabbit IgG	Cy3	Life Technologies (reference no. A10520)	1:800	#1
Goat polyclonal IgG anti-rabbit IgG	*Alexa Fluor 488*	Life Technologies (reference no. A11034)	1:500	#2
Donkey polyclonal IgG anti-rabbit IgG	*Alexa Fluor 488*	Abcam (reference no. ab150073)	1:800	#3
Goat polyclonal IgG anti-chicken IgG	Cy3	Abcam (reference no. ab97145)	1:800	#4

treatment. Sections were labeled with a probe for Cre (Cre-O4, reference no. 546951, from Advanced Cell Diagnostics). Probes were visualized with Opal fluorophore (OpalTM 570, reference no. FP1488001KT, Akoya biosciences). Sections were counterstained with DAPI and mounted with ProLongTM Gold antifade (reference no. P36935, Invitrogen).

Slides were imaged with a Nikon Ti/CSU-W1 spinning disc confocal microscope equipped with 405- and 561-nm excitation lasers using a 60× objective. Cells of interest were identified on the basis of the expression of *Cre*. In VNO sections of *Fpr-rs3*$^{iVenus\text{-}iCre}$ mice, cells within the medial region of the VNO were excluded from the analysis. Images of eight z stacks covering a range of 5 μm were acquired and exported as orthogonal maximum intensity projections for analysis. The same acquisition parameters were used for all images. Images were analyzed with Image J (version 1.53q). Maximum projection images including only the *Cre* channel (pseudo-colored in green) were analyzed by drawing a region of interest (ROI) around the cell with a constant area of 710 μm^2 (including the entire VSN cell body and nearest background) and measuring the mean fluorescence intensity within this ROI.

Microscopy

Population analyses of transgene-expressing cells

For the identification and the quantification of Fpr transgene-expressing cells, images were acquired with a Leica DM5500 epifluorescence microscope using the 20× objective centered around one VNO side. Cells were counted and labeled manually on the acquired images.

Whole-mount pictures of AOBs

AOBs were exposed on their dorsal side, immerged in PBS, and imaged with a Zeiss LSM 700 confocal microscope with the 10× objective.

Statistics and reproducibility

Statistical tests and data representation were computed in R version 4. All data points are represented in the graphs. When applicable, dispersion is shown in the background with violin plots, boxplots, or barplots and error bars showing means and SDs, respectively. All comparisons performed were two-tailed. Continuity correction was applied to Wilcoxon rank sum tests. Yates' continuity correction was applied to chi-squared tests. Significance was assessed for $P < 0.05$. All biological replicates were obtained from age- and sex-matched animals. Analyses were performed without blinding of the investigator. No data were excluded from the analyses.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abn7450

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. I. Rodriguez, U. Boehm, Pheromone sensing in mice. *Results Probl. Cell Differ.* **47**, 77–96 (2009).
2. K. Del Punta, T. Leinders-Zufall, I. Rodriguez, D. Jukam, C. J. Wysocki, S. Ogawa, F. Zufall, P. Mombaerts, Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature* **419**, 70–74 (2002).
3. Y. Isogai, S. Si, L. Pont-Lezica, T. Tan, V. Kapoor, V. N. Murthy, C. Dulac, Molecular organization of vomeronasal chemoreception. *Nature* **478**, 241–245 (2011).
4. I. Rodriguez, Vomeronasal receptors: V1Rs, V2Rs and FPRs, in *Chemosensory Transduction: The Detection of Odors, Tastes, and other Chemostimuli*, F. Zufall, S. D. Munger, Ed. (Elsevier Inc., Academic Press, 2016), pp. 175–190.
5. J. M. Young, H. F. Massa, L. Hsu, B. J. Trask, Extreme variability among mammalian V1R gene families. *Genome Res.* **20**, 10–18 (2010).
6. S. Riviere, L. Challet, D. Fluegge, M. Spehr, I. Rodriguez, Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature* **459**, 574–577 (2009).
7. S. D. Liberles, L. F. Horowitz, D. Kuang, J. J. Contos, K. L. Wilson, J. Siltberg-Liberles, D. A. Liberles, L. B. Buck, Formyl peptide receptors are candidate chemosensory receptors in the vomeronasal organ. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9842–9847 (2009).
8. Q. Dietschi, J. Tuberosa, L. Rösingh, G. Loichot, M. Ruedi, A. Carleton, I. Rodriguez, Evolution of immune chemoreceptors into sensors of the outside world. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7397–7402 (2017).
9. I. Rodriguez, P. Feinstein, P. Mombaerts, Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. *Cell* **97**, 199–208 (1999).
10. L. Belluscio, G. Koentges, R. Axel, C. Dulac, A map of pheromone receptor activation in the mammalian brain. *Cell* **97**, 209–220 (1999).
11. Q. Dietschi, A. Assens, L. Challet, A. Carleton, I. Rodriguez, Convergence of FPR-rs3-expressing neurons in the mouse accessory olfactory bulb. *Mol. Cell. Neurosci.* **56**, 140–147 (2013).
12. S. Wagner, A. L. Gresser, A. T. Torello, C. Dulac, A multireceptor genetic approach uncovers an ordered integration of VNO sensory inputs in the accessory olfactory bulb. *Neuron* **50**, 697–709 (2006).
13. D. Roppolo, S. Vollery, C. D. Kan, C. Lüscher, M. C. Broillet, I. Rodriguez, Gene cluster lock after pheromone receptor gene choice. *EMBO J.* **26**, 3423–3430 (2007).
14. L. Capello, D. Roppolo, V. P. Jungo, P. Feinstein, I. Rodriguez, A common gene exclusion mechanism used by two chemosensory systems. *Eur. J. Neurosci.* **29**, 671–678 (2009).
15. R. P. Lane, J. Young, T. Newman, B. J. Trask, Species specificity in rodent pheromone receptor repertoires. *Genome Res.* **14**, 603–608 (2004).
16. R. P. Lane, T. Cutforth, R. Axel, L. Hood, B. J. Trask, Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 291–296 (2002).
17. A. D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. M. Dowall, A. M. Mahon, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, G. IIsley, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino, P. Flicek, Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
18. A. L. Mitchell, T. K. Attwood, P. C. Babbitt, M. Blum, P. Bork, A. Bridge, S. D. Brown, H. Y. Chang, S. el-Gebali, M. I. Fraser, J. Gough, D. R. Haft, H. Huang, I. Letunic, R. Lopez, A. Luciani, F. Madeira, A. Marchler-Bauer, H. Mi, D. A. Natale, M. Necci, G. Nuka, C. Orengo, A. P. Pandurangan, T. Paysan-Lafosse, S. Pesseat, S. C. Potter, M. A. Qureshi, N. D. Rawlings, N. Redaschi, L. J. Richardson, C. Rivoire, G. A. Salazar, A. Sangrador-Vegas, C. J. A. Sigrist, I. Sillitoe, G. G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, S. Y. Yong, R. D. Finn, InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).
19. K. Katoh, J. Rozewicki, K. D. Yamada, MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).
20. G. F. Jenks, The data model concept in statistical mapping. *Int. J. Cartogr.* **7**, 186–190 (1967).
21. R. W. Meredith, J. E. Janecka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik, T. L. L. Simao, T. Stadler, D. L. Rabosky, R. L. Honeycutt, J. J. Flynn, C. M. Ingram, C. Steiner, T. L. Williams, T. J. Robinson, A. Burk-Herrick, M. Westerman, N. A. Ayoub, M. S. Springer, W. J. Murphy, Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
22. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2-Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).
23. L. Fodoulian, O. Gschwend, C. Huber, S. Mutel, R. F. Salazar, R. Leone, J.-R. Renfer, K. Ekundayo, I. Rodriguez, A. Carleton, The claustrum-medial prefrontal cortex network controls attentional set-shifting. *bioRxiv*, 2020.10.14.339259 (2020).
24. A. Saxena, A. Wagatsuma, Y. Noro, T. Kuji, A. Asaka-Oba, A. Watahiki, C. Gurnot, M. Fagiolini, T. K. Hensch, P. Carninci, Trehalose-enhanced isolation of neuronal sub-types from adult mouse brain. *Biotechniques* **52**, 381–385 (2012).
25. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
26. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

27. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
28. T. Smith, A. Heger, I. Sudbery, UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
29. C. Mayer, C. Hafemeister, R. C. Bandler, R. Machold, R. B. Brito, X. Jaglin, K. Allaway, A. Butler, G. Fishell, R. Satija, Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
30. H. Luche, O. Weber, T. Nageswara Rao, C. Blum, H. J. Fehling, Faithful activation of an extra-bright red fluorescent protein in "knock-in" Cre-reporter mice ideally suited for lineage tracing studies. *Eur. J. Immunol.* **37**, 43–53 (2007).

Acknowledgments: We thank J. Dal Col and F. Resende for expert technical assistance and the iGE3 Genomics Platform at the University of Geneva for assistance during bulk and scRNA-seq experiments. **Funding:** This research was supported by the University of Geneva and the Swiss National Science Foundation (grant numbers 31003A_172878 to A.C. and 310030_189153 to I.R.). **Author contributions:** I.R. and A.C. conceived the project, acquired funding, supervised research, and wrote the manuscript. Q.D., J.T., L.F., and M.B. participated to the experimental design, performed experiments, collected data, and interpreted results. P.F. provided resources. C.K., J.C., and V.P. participated to the data collection and interpretation of the results. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Bulk and scRNA-seq data have been deposited in NCBI GEO (Gene Expression Omnibus; www.ncbi.nlm.nih.gov/geo/), can be accessed using the superseries GSE190613 accession number, and are publicly available as of the date of publication. Coordinates of V1r intact CDS and pseudogenic sequences, VSN counts from the scRNA-seq data, and custom genome annotation files for the mouse and the rat are available in the Supplementary Materials. All other data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 17 December 2021
Accepted 15 September 2022
Published 16 November 2022
10.1126/sciadv.abn7450