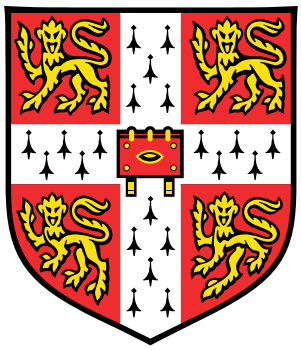
Genetic and environmental regulation of olfactory sensory neurone diversity



Ximena Ibarra Soria

Wellcome Trust Sanger Institute

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Wolfson College

September 2015

To mom and James for being there every step of the way.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation contains fewer than 60,000 words excluding bibliography, figures and appendices, as per the requirements of the Degree Committee for the Faculty of Biology.

Ximena Ibarra Soria
September 2015

Acknowledgements

First and foremost, I would like to thank my supervisor, Darren Logan, for his continued support in and outside the lab. For always being available and willing to help; for his patience and encouragement. Very importantly, for teaching me how to think about science and for sharing all his wisdom and wit. There are a few people that have had a tremendous impact in my academic formation and, without a doubt, Darren is one of them; for that I will always be thankful.

I would also like to acknowledge all the people in the lab. Gabi who provided great advice and ideas to improve experiments, and was always happy to help. Luis, for being open to collaborate and combine the best of our abilities to create greater things. Maria, for working alongside me and making the lab a happier place. Elizabeth, Cristina and Sebastian for all their help with the wet lab. Laura and Sophia for useful discussions.

A big thank you to team 113, who made me an honorary member of their group. For all the lunch-time *insightful* discussions and laughter; and for including me in their dinners, Christmas celebrations, trips and outings. You have made my PhD a much happier one. Special thanks go to Daniela, who has always been there for me, sharing every step of the way. For her friendship and love, for the long talks over brunch, for the encouragement on the tough times and for always making sure I was ok.

I would also like to thank John Marioni, my secondary supervisor, for his help with the bioinformatics and for letting me pick his brains when I was struggling to understand my data. And to the rest of my thesis committee, David Adams and Gregory Jefferis who, along with John and Darren, provided very useful advice and comments on the progress of my research and how to make it better.

A lot of my PhD work was greatly facilitated by the staff of the Research Scientific Facility and the Illumina Bespoke Sequencing team. Special thanks to Mairi Kusma, Andrea Kirton and Nathalie Smerdon for invaluable help with many projects.

I owe big acknowledgment to my parents, for making sure I had the best opportunities to achieve whatever I wanted. To mom, for giving up so much so that I could have the best life possible; without your support I wouldn't be here. To my brother, for drawing

pretty mice I could use without infringing copyright, and for his support and advice along the way. Also, to Yuyi, who has always kept an eye on me even while being so far away.

A big thank you to James, for being along my side during this journey. For sharing his life with me and picking me up in the toughest times. For all his love and for making me a happier, better person.

Finally, I would like to acknowledge the Wellcome Trust for their generous support.

Abstract

Animals use their sense of smell to gather plethora of information about their surroundings. The detection of odorants occurs in the main olfactory epithelium (MOE), which contains olfactory sensory neurones (OSNs) among other cell types; these express olfactory receptors (ORs) that bind to odorants. Each OSN expresses only one allele of one OR gene from a family of over 1,200 in the mouse genome. Thus, the mouse nose has over 1,200 different OSN types, each characterised by the OR expressed. High levels of genomic variation have been reported both in the mouse and human OR repertoire. This is thought to contribute to the unique sense of smell each individual has, but a large proportion of the observed phenotypic variance remains unaccounted for.

In this dissertation, I present the results from an RNAseq-based approach used to quantify the OSN repertoire of the mouse. Firstly, I validated the accuracy and reproducibility of this technology to study the olfactory system. I then characterised the transcriptome of the MOE and of the OSNs as a population and at the single-cell level. This allowed me to conclusively prove that OR expression is indeed monogenic and monoallelic. Then, I demonstrated that the method is sensitive enough to detect the expression of almost the complete OR repertoire. Also, I was able to annotate full-length gene models for many OR genes.

Secondly, I explored the diversity of OSN types in three inbred strains of mice (C57BL/6, CAST/EiJ and 129S5) via their OR gene expression levels. I found that each strain has a unique and reproducible distribution of OSNs in their noses, and that genomic variation instructs this neuronal variance in *cis*. Finally, I analysed the plasticity of the distribution of the different classes of OSNs by stimulating animals with particular odorants. Exposure to an enriched olfactory environment results in the differential expression of dozens of OR genes in a reproducible and specific manner. These changes increase with time and are reversible. These data allow to comprehensively explore and dissect the effects of genetic and environmental variation on the regulation of OR expression and OSN repertoire. Together they generate an olfactory sensory system that is individually unique.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Some of the material presented in this chapter has been previously published in reference [1]. I confirm the sections used are my own work and are reproduced with kind permission from Springer Science and Business Media.

1.1 The mammalian olfactory system.

Animals live in a constantly changing and ever challenging world. For many, the olfactory sensory system is fundamental to accomplish tasks essential for survival and reproduction; from finding food to identifying if it's spoiled, from detecting predators to natural dangers like fires, from identifying conspecifics to determining if they are suitable for mating[2]. All these processes are guided by olfactory cues, which become paramount in nocturnal animals or those with a less developed visual and auditory systems. The appropriate detection and correct interpretation of such cues is essential to eliciting an adequate response. Most mammals have developed a complex olfactory system, composed of several organs specialised in the detection of a plethora of chemosignals. The two main components are the main olfactory epithelium (MOE) and the vomeronasal organ (VNO), but the Grueneberg ganglion (GG) and the septal organ (SO) also play a role in olfaction[3] (Figure 1.1).

Traditionally, it has been considered that the MOE is involved in recognising volatile common odorants; these are low molecular weight molecules that can be perceived as odorous via the olfactory system[5]. The number of different odorants that exist is still a debated question; figures range from thousands to hundreds of thousands or even

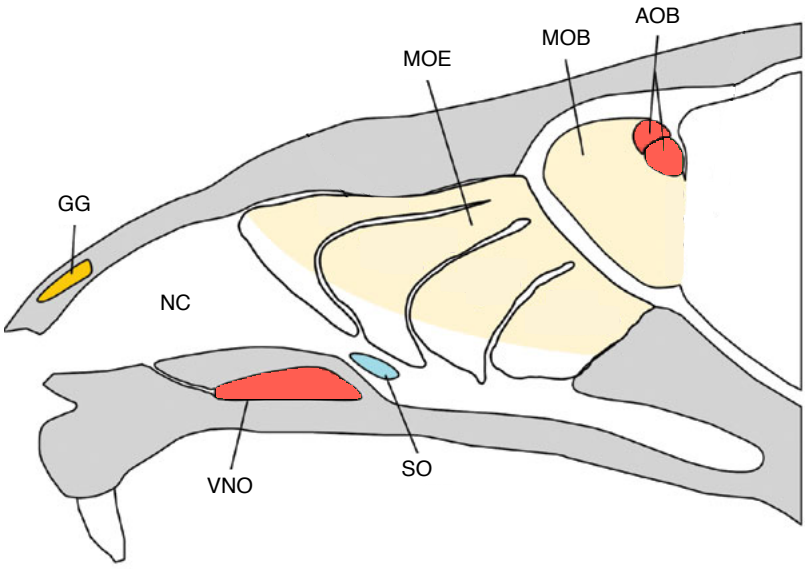


Figure 1.1 – The mammalian olfactory system. Schematic of the mouse nose (sagittal view) and the different components of the olfactory system. At the back of the nasal cavity (NC) is the main olfactory epithelium (MOE). Directly above the roof of the mouth, at the base of the nasal cavity, is the vomeronasal organ (VNO). Near the ventral end of the nasal septum at the entrance of the nasopharynx is the septal organ (SO), which is separated from the VNO and MOE by respiratory epithelium. Finally, at the rostral end of the nasal cavity, just inside the nostrils, is the Gruenberg ganglion (GG). The sensory neurones from the MOE, SO and GG project axons to the main olfactory bulb (MOB), while the neurones from the VNO project to the accessory olfactory bulb (AOB). Figure adapted by permission from Macmillan Publishers Ltd: Nature ([4]), copyright (2006).

millions[2, 6]. Similarly, the number of different odorants that animals can detect is unknown. This is a daunting question since the odorant space is not defined. The molecules that can be odorous occupy a broad range of physicochemical properties, shapes and sizes; they can be produced by living organisms or can be inorganic substances[5]. More often than not, stimuli are present as mixtures, with varying concentrations of each component; interactions between different odorants and their proportions all have an impact on the ability of the olfactory system to detect and interpret them as a smell. It is also important to differentiate between detection and discrimination; an organism can be capable of detecting two different odorants, but they may *smell* the same. A recent paper claimed that humans are able to discriminate at least 1 trillion odorants, based on calculations of how many mixtures of 30 different odorants could be discriminated by a set of individuals[7]. However, the statistical framework that led to this calculation has been challenged[8, 9] leaving the question of how many odorants can be discriminated (or detected) still unanswered.

On the other hand, the VNO has been considered to specialise in the detection of pheromones. A pheromone is typically defined as a social cue that is transmitted between two animals of the same species[10]; it usually induces a particular behaviour or

endocrine change on the receiver individual. These compounds do not need to be volatile and are often relatively large organic molecules, peptides and proteins[5]. Pheromones are used for social interaction and communication; animals obtain diverse information such as sex, strain, health status and reproductive state from pheromonal signals[5, 11]. Upon detection, some of these cues result in behaviours such as aggression or mating, or can have lasting physiological effects such as puberty acceleration in females[11].

In recent years it has become apparent that the separation between the MOE and VNO as sole detectors of odorants and pheromones respectively is not as definite as proposed before. A growing body of work has now documented several examples in which the same olfactory cues are detected by both the VNO and MOE, utilising different detection and signalling mechanisms, and generating different responses. The picture emerging suggests that all the different components of the olfactory system work together to sense the chemical world, and their respective signals are integrated in cortical areas of the brain[12].

1.1.1 The main olfactory epithelium.

The main olfactory epithelium (MOE) is located at the back of the nasal cavity. It is a pseudostratified columnar epithelium composed of three primary cell types: the olfactory sensory neurones (OSNs), sustentacular or supporting cells and basal cells (Figure 1.2).

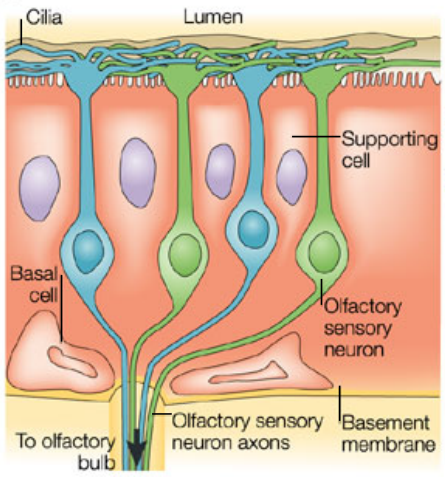


Figure 1.2 – Composition of the MOE. Schematic representing the different cell types present in the main olfactory epithelium and their organisation. In blue and green are the bipolar olfactory sensory neurones (OSNs), that extend dendrites towards the lumen of the MOE; dendrites terminate in long cilia that sit on the surface and interact with odorants. OSNs send one axon each towards the main olfactory bulb. OSNs are surrounded by supporting cells, which span the whole thickness of the epithelium. At the basement membrane can be found basal cells, that have the ability to proliferate and differentiate into new OSNs. Figure adapted by permission from Macmillan Publishers Ltd: Nature Reviews Neuroscience ([2]), copyright (2004).

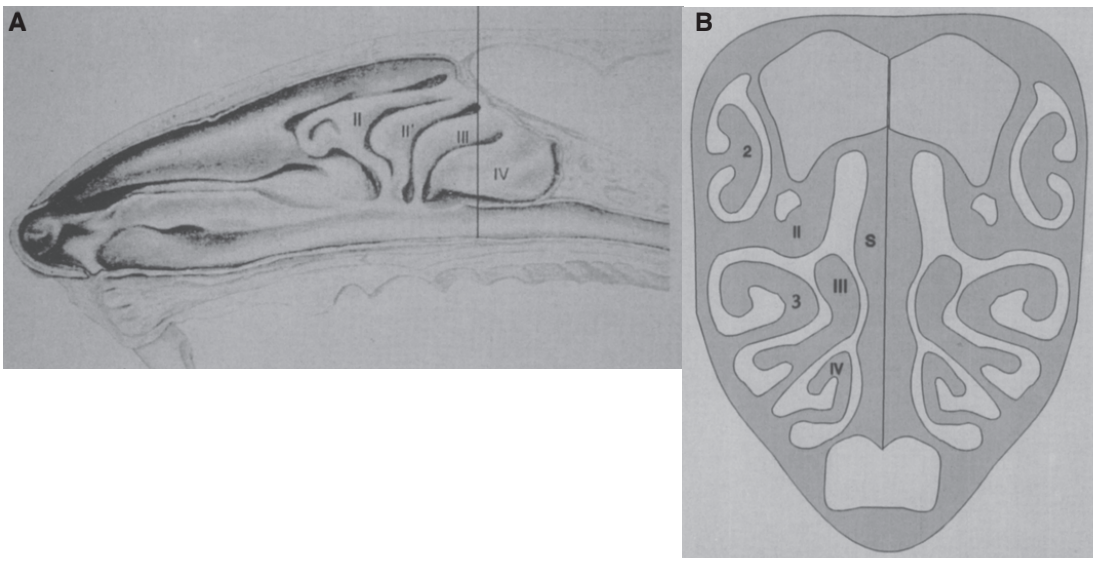


Figure 1.3 – **Turbinate structure of the MOE. A)** Sagittal view of the mouse nose. The vertical line indicates the plane of section represented in **B)** in a coronal view. The numerals indicate the individual turbinates. II, II', III and IV endoturbinates; 2, 3 ectoturbinates. s, nasal septum. Reproduced from ([18]) with kind permission from Springer Science and Business Media.

It sits on top of connective tissue lamina propria which contains glands and blood vessels [13] and is shaped by surrounding cartilage that forms a number of outcroppings, called turbinates[6] (Figure 1.3). The bipolar OSNs are the basic units of olfactory detection. From their apical side extends a single dendrite that protrudes into the epithelial surface; such dendrites terminate in an enlargement, called the olfactory vesicle or knob, which contains numerous cilia. Each olfactory vesicle contains around 5 to 25 cilia, which can be as long as twice the length of the OSN itself (so much so, that these were initially called *'olfactory hairs'*[14]); these all lie on the surface of the epithelium, which is covered by an aqueous mucus layer where chemical molecules dissolve and are able to interact with the cilia. This is the primary site of odorant detection[13–16]. From the basal region of the OSN extends an unbranched axon that projects to the main olfactory bulb (MOB) in the forebrain; axons pass through the lamina propria in large bundles which then travel through the cribriform plate to reach the MOB[6, 13, 16, 17].

OSNs are embedded within supporting cells, which are columnar epithelial cells that cover the whole thickness of the epithelium; they resemble glia and have microvilli on their apical side. These cells originate from non-nervous ectoderm. Their principal functions are to protect and provide support for the OSNs, but are also involved in secreting some of the mucus components and in phagocytosis of degenerated OSNs and dendritic fragments [13, 16, 17]. Finally, the basal cells are neuroblasts found near the basal lamina, and constitute the stem cells of the olfactory neuroepithelium. These are

further subdivided into horizontal and globose basal cells (HBCs and GBCs respectively). HBCs lie in a single layer in direct contact with the basal lamina; they are triangular in shape and express cytokeratin[19]. These cells rarely proliferate *in vivo* and in normal conditions stay quiescent; however they have the ability to produce both neurones and glia *in vitro* and *in vivo* when the olfactory epithelium is severely injured[20]. GBCs are found on top of HBCs; they are spherical and express the neural adhesion protein N-CAM. Tracing studies have shown the GBCs are proliferative and can differentiate into OSNs[19, 21]. Upon injury affecting only the OSNs, the GBCs proliferate to repopulate the neuroepithelium, and the HBCs remain quiescent[20]. This proliferative capacity is maintained throughout the animal's life span[13, 15, 19], which is of fundamental importance given the direct proximity of OSNs with the environment. Harmful chemicals and pathogens constantly reach the neurones, making the ability to regenerate damaged cells paramount[22].

The differentiation process of GBCs into OSNs is accompanied by migration of the cells from the basal to the apical part of the epithelium; fully mature OSNs express the olfactory marker protein (OMP) and are found most apically, whereas immature OSNs are intermediate between GBCs and mature OSNs[19]. As a pseudostratified tissue, the OSNs are organised in up to 8 layers in the thickest regions of the epithelium[23]; the somata of the sustentacular cells sit on top of the OSN layer.

The lamina propria provides structural support for the olfactory epithelium; it contains Schwann cells that surround the OSN axons as they exit towards the MOB. Another important component found in the lamina propria are the Bowman's glands. These extend a single duct through the epithelium into the mucosal surface and are instrumental in the production of mucus[13]. One of the main functions of the mucus is to protect the epithelium from drying out. Also, it contains several enzymes that help combat infection, and odorant binding proteins that aid in the transport and stabilisation of ligands[24].

The OSNs in the MOE project their axons to the MOB, where they synapse with the dendrites of mitral and tufted cells; this is the first relay station in olfactory processing. Thousands of OSNs synapse to only 5 to 25 mitral cells[6]. In turn, both mitral and tufted cells project to the olfactory cortex through the lateral olfactory tract, forming two parallel projection networks[25]. The principal regions in the olfactory cortex that receive inputs from the MOB include the anterior olfactory nucleus, taenia tecta, olfactory tubercle, piriform cortex, nucleus of lateral olfactory tract, anterior cortical amygdaloid nucleus, posterolateral cortical amygdaloid nucleus, and entorhinal cortex[5, 25]. Further

signalling from the olfactory cortex proceeds to cortical areas such as the orbitofrontal cortex, the amygdala and the medial preoptic area in the hypothalamus[5].

Regeneration of the OSN population.

After birth, the MOE continues to develop with rapid growth in the first postnatal weeks. In adults, however, both the volume and surface of the epithelium remains constant[22]. In 8-week old mice, it has been estimated that the MOE contains around 10 million mature OSNs –as determined by OMP expression– with no significant differences for male and female individuals[23]. In mice, growth stops after three months of age[26] but OSNs continue to be produced over the life-span of the animal; this implies a tight regulation of turn-over of OSNs with balanced neurogenesis and apoptosis rates[22].

The life-span of individual OSNs has been assessed by labelling with thymidine analogs that mark proliferating cells. In initial studies it was observed that labelled cells had disappeared after 30 days post-labelling[27]; this time-frame is widely cited in the literature. However, further analyses found that some labelled neurones were still present up to 3[28], or even 12 months after injection of the analogs[26]. A different approach consisting on retrograde labelling of neurones by injecting the MOB, also found marked OSNs up to 3 months after labelling[29]. Therefore, there is evidence that some OSNs can live for several months. Nonetheless, all these studies also agreed that a great proportion of the marked cells disappeared between 14 and 30 days[28, 30] and only a small subset survived for longer periods.

The regenerative capacity of the MOE has also been studied upon injury. Several methods have been developed to eliminate the OSN population, based on treatment with toxic chemicals (such as methyl bromide or zinc sulphate) or by olfactory nerve axotomy or bulbectomy; all of these result in the death of the OSNs and chemical methods usually affect other cell types as well[31, 32]. In all cases, extensive neurogenesis is observed, with the basal progenitor cells replenishing the neuronal population of the epithelium. This is accompanied by a recovery of olfactory function. The degree of the recovery usually depends on how severe the injury is and the age of the animal[31–33].

Olfactory signalling.

Olfaction is initiated by the recognition of odorous molecules by the OSNs; this is achieved by olfactory receptors (ORs) expressed by OSNs and localised to the membranes of their cilia. When an OR binds its ligand, a signalling cascade is activated to

produce an action potential that travels into the MOB, where the information is processed. The different components involved in this transduction pathway were identified in the 1980's, exploiting the knowledge from the signalling mechanisms employed by the visual system. Firstly, it was shown than an adenylate cyclase is highly enriched in the cilia from the frog olfactory neuroepithelium, compared to the brain or total MOE. Furthermore, the activity of the enzyme was shown to be dose-dependent when the membranes were stimulated with odorants. This supported the notion that the protein is specifically involved in olfactory-mediated signal transduction. Additionally, activity was only observed in the presence of GTP, which suggested that the coupling between the cyclase and the receptors was occurring via a guanine nucleotide binding protein (G-protein)[34]. These findings were later replicated in rat olfactory cilia[35]. Adenylate cyclase is an enzyme that converts ATP into cAMP, which in turns serves as a second messenger in the transduction cascade; therefore, when it is activated by odorants, there is a rapid, dose-dependent increase in the levels of cAMP. Importantly, the increase occurs well before the membrane is depolarised to induce an action potential, supporting it's role as a second messenger[36, 37]; cAMP in turn activates a cyclic nucleotide-gated (CNG) channel that allows the change in the membrane's potential.

Using patch-clamp in the cilia membranes, it was indeed confirmed that a rise in cAMP resulted in an increase in the membrane conductance[38]. The responsible gene for this effect was cloned revealing a protein with 57% identity to the cGMP-gated channel expressed in bovine rods; the C-terminal domain, where the cyclic nucleotide binding site resides, is highly conserved between the two. This protein was shown to be expressed specifically in the MOE and, what's more, particularly in the OSNs[39]. Further studies on the electrophisiology of OSN cilia revealed that the activation of this CNG channel by elevation of cAMP results in an influx of Ca^{2+}; there are, in turn, Ca^{2+}-dependent Cl^{-}channels that allow efflux of this ion, further amplifying the inward current and boosting the signal above basal noise[40–43].

Lastly, the dependance of the adenylate cyclase on GTP ignited the search for an olfactory specific G-protein; these are well known for their role in coupling membrane-bound receptors to second-messenger enzymes or ion channels. By hybridisation with a degenerate probe from a highly conserved GTP-binding domain, a novel G-protein was obtained from an MOE cDNA library. It was identified as a close homolog (88% identity) of another G-protein expressed in the MOE ($G\alpha_s$) but that is predominantly present in non-neuronal cells. This new protein was shown to be specific to the OSNs in the MOE and was therefore named $G\alpha_{olf}$. Finally, it was demonstrated that it was capable

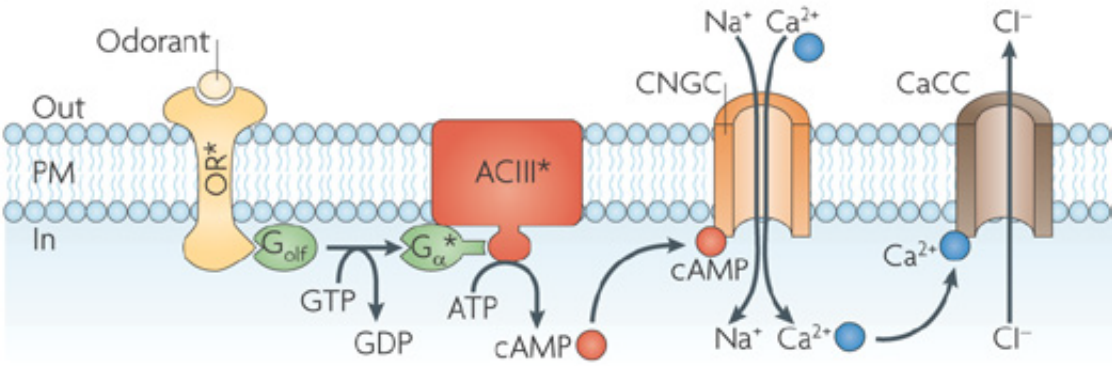


Figure 1.4 – Olfactory signal transduction cascade. When an odorant binds to an olfactory receptor (OR) it activates the trimeric G protein $G\alpha_{olf}$ which in turns activates the adenylate cyclase ACIII. ACIII catalyses the conversion of ATP into cAMP, which serves as a second messenger to open cyclic nucleotide gated channels (CNGC) that allow the influx of sodium and calcium. In turn, calcium activates calcium-dependent chloride channels (CaCC) that allow the efflux of this ion. The movement of ions results in the depolarisation of the membrane and the generation of an action potential. PM, plasma membrane. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Neuroscience ([47]), copyright (2010).

of stimulating adenylate cyclase activity[44]. All together, these components assemble into a transduction signalling pathway whereby an olfactory receptor is activated by binding to its ligand; this in turns stimulates $G\alpha_{olf}$ (*Gnal* in mice) which can activate the adenylate cyclase (*Adcy3* in mice[45]); production of cAMP then acts upon a CNG channel present in the plasma membrane (*Cnga2, Cnga4* and *Cnga1* in mice) which results in Ca^{2+}influx which in turns activates Ca^{2+}-dependent Cl^-channels (*Ano2* in mice[46]); flux of ions through both channels induce an alteration of the membrane's potential and, ultimately, lead to the generation of an action potential that can travel through the OSN's axon into the brain (Figure 1.4).

All the major components on this transduction cascade have been individually knocked out in mice, to reveal their indispensable function in olfactory-mediated signalling. Knockout (KO) of *Adcy3*, *Cnga2* and *Gnal* all result in animals that cannot smell (anosmic). Most homozygotes die within two days after being born because they fail to suckle[45, 48–50], a process that has been shown to depend on olfactory cues[51]. By reducing the litter sizes and eliminating the competition from wild-type littermates, up to 10% of the KO animals manage to survive to adulthood. Interestingly, knocking out any given gene doesn't seem to have an appreciable effect on the anatomy of the MOE or the expression of the other genes in the signalling pathway. *Omp* expression seems normal in most cases as does OR gene expression[45, 50] with the exception of the *Cnga2* KO, which has a significantly reduced number of *Omp* expressing mature OSNs, a considerably lower number of immature neurones and a smaller MOB[48, 49]. Electroolfactogram

(EOG) recordings measure the extracellular field potential that results from activation of OSNs in response to odorants; the measurements are a summation across all the cells around the recording electrode. EOG recordings revealed no olfactory-mediated activity in the OSNs lacking any of the signalling components, upon stimulation with a variety of odorants[45, 48–50] and even biological substances such as urine[48]. Furthermore, the *Adcy3* KO animals were shown to be anosmic by behavioural tests, where homozygote animals failed to associate an odour cue with either an aversive or positive cue[45]. Therefore, all the results indicate that animals that lack any of *Adcy3*, *Cnga2* or *Gnal* are largely unable to smell.

The olfactory receptor genes.

After identifying the different components involved in olfactory signalling, the piece still missing was the OR itself. OR genes were initially identified by Linda Buck and Richard Axel in 1991, under the assumption that the receptor genes should be able to transduce intracellular signals by coupling to Gα_{olf}. Additionally, given the myriad odorants that animals can identify, receptors would be most likely part of a multi-gene family that should be expressed in the MOE. With these premises in mind, Buck created degenerate PCR primers based on the sequences for G-protein coupled receptors (GPCRs) known at the time, and used them to amplify homologous sequences from cDNA from the olfactory epithelium. The obtained products were further analysed to identify those that contained several different sequences, as would be the case for a multi-gene family. One of the PCR products had these characteristics and sequencing of individual clones revealed that, indeed, it contained different DNA sequences that shared common motifs[52].

Further analysis demonstrated that the identified genes were a novel class of GPCRs, with the characteristic seven transmembrane domain, connected by intra- and extracellular loops of different lengths; their N-termini is located on the extracellular side of the plasma membrane while the C-termini is in the cytoplasm[6, 52]. Most OR genes have two to five exons but, similar to other GPCRs, they have their coding sequence (CDS) contained within a single exon[6, 53]. Subsequently, many other OR genes were identified in several species. However, it wasn't until the advent of whole genome sequencing that the complete repertoires of OR genes were characterised. Availability of genome drafts allowed the computational prediction of many more of these genes and it soon became evident that they represent the biggest multi-gene family in the mammalian genome[54–58]. In the mouse, there are 1250 annotated OR genes (named *Olfr*), 15% of which are classified as pseudogenes. They are dispersed along the genome, occupying

most chromosomes, and accommodated in tight clusters of varying sizes that range from one to several hundred genes. The exact number of clusters depends on the definition used, but roughly represent 40 to 50 different loci; genes within a cluster tend to be separated by an average of 21 kb though this varies greatly. Most of the big clusters contain non-OR genes interspersed with the receptors[55–58].

Phylogenetic analyses have revealed that the OR repertoire is composed of two distinct types of genes, named class I and class II receptors[59]. The class I genes account for 10% of the total number of ORs and are more closely related to fish OR genes[55, 56, 59]; in the mouse, they are all found in one big cluster in chromosome 7[55]. On the other hand, class II ORs are specific to terrestrial vertebrates. The OR repertoire has been subdivided into families by grouping all those receptors that share at least 40% identity at the amino acid level. This cutoff was chosen because any given OR shares at least 40% identity with its nearest neighbour, but at most 38% identity to any of the other GPCRs[54]. For the mouse, ORs are grouped into 149 families, 29 of which contain class I receptors; the number of ORs per family varies from one to 97 different genes[58] (Figure 1.5). Over half of the ORs have at least one paralog with more than 80% identity and some genes can be nearly identical. However, the diversity between genes of different families is very large[2, 58], with an average identity of 37% that can drop as low as 18%[60]. Genes that are closely related tend to be found in the same locus, which suggests that the expansion of the OR gene repertoire has occurred by local events of gene duplication followed by diversification[54–57, 61].

The sequences of putatively functional OR genes contain conserved motifs in different regions that are shared by a large proportion of the receptors, even across species, and that differentiate them from other types of GPCRs. The transmembrane (TM) helices 4 and 5, and part of the TM3 are highly variable as are the end of the N- and C-termini[63, 64]. It has been proposed that these three TM domains face each other in the plasma membrane, creating a pocket that is probably the site of ligand binding[6, 63]. The rest of the TMs and the intracellular loops are more conserved and have characteristic motifs that are sufficient to characterise a seven TM protein as an OR. Additional motifs are present in subsets of ORs; for example, a few motifs are specific to class I or class II ORs, generally occupying the extracellular side of TM6[58, 64]. Furthermore, specific combinations of motifs have been identified in groups of ORs that interact with related ligands and might therefore be important for ligand recognition[64].

Additionally, several residues are well conserved in most OR sequences and have been proposed as key amino acids in the structure and stabilisation of the receptor proteins;

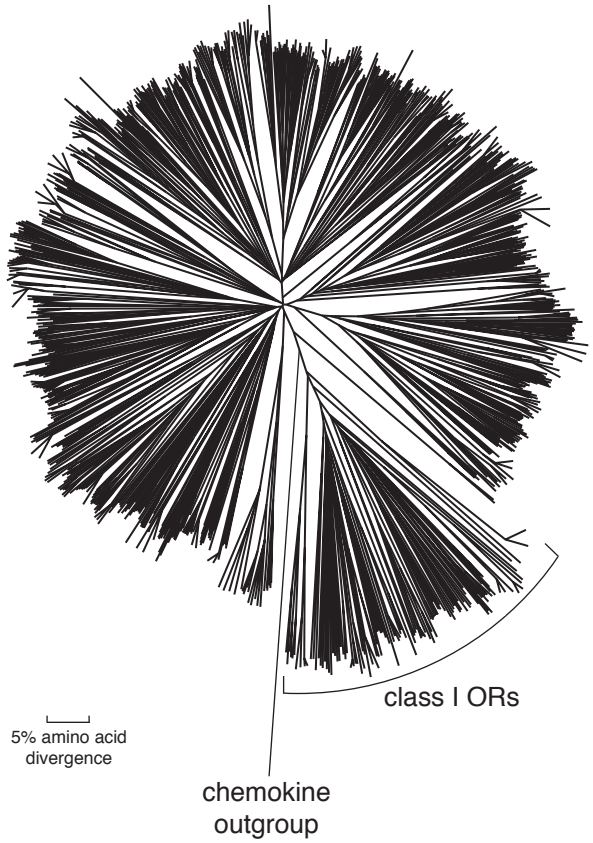


Figure 1.5 – The mouse olfactory receptor gene family. Phylogenetic tree of the mouse olfactory receptor genes. The class I genes are indicated and the remaining are class II. Reproduced from [62].

these tend to be conserved in other families of GPCRs also. The most prominent are cysteine residues found in the extracellular loops 2 and 3 as well as in TM3, which could form disulphide bonds[64, 65], and there are also some potential glycosylation sites which could be important in the regulation of expression or the stability of the protein[64].

OR expression is monogenic and monoallelic.

Further studies on the expression of several OR genes in the MOE of rodents revealed that different subfamilies of receptors are expressed in non-overlapping subsets of OSNs. The initial experiments used *in situ* hybridisation to explore the pattern of expression of ORs; the probes utilised recognised up to 20 related OR genes, all from the same subfamily. With these, a subset of cells were labelled, in a punctate pattern, scattered throughout a region of the epithelium[18, 66, 67]. The labelled cells were within the layers occupied by OSNs and no signal was detectable in the basal or sustentacular cell layers[18]. Each labelled OSN was surrounded by many others that did not express

the same OR genes. This implied that each OSN expresses only a subset of receptors. Additionally, when all the different probes were hybridised together, the number of labelled neurones was very close to the sum of cells labelled with each individual probe, suggesting that the genes recognised by each probe were expressed in distinct subsets of cells[18, 66–68]. On average, each receptor has been observed in only 0.1% of the neuronal population, suggesting that each OSN expresses a single, or very few OR genes[66, 67].

In a further study, OR expression was assessed by single-cell RT-PCR. Using degenerate primers, PCR products were amplified from cDNA obtained from isolated OSNs from the dorsal region of the epithelium. The PCR was successful in 18 out of 26 tested cells and in all cases the product represented a single OR gene. This further supported the idea that each OSN expresses only a subset of the OR repertoire and, what's more, it suggested that a single receptor was present in each neurone[69]. This notion was further supported by studies using transgenic mouse lines, carrying several receptors tagged with different reporter genes; the expression of each reporter could be observed in a particular subset of OSNs that was mutually exclusive with the population labelled by other reporter genes[70].

Even more remarkable was the finding that monogenic expression extends to transgenes. A yeast artificial chromosome (YAC) carrying three mouse OR genes, one of which was tagged with lacZ, was inserted at random locations in the genome; additionally, the corresponding endogenous OR gene was tagged with GFP. When animals were stained for both reporters, each was expressed in a distinct group of OSNs, with very few cells showing co-staining. Moreover, when differentially tagged transgenes were inserted to produce transgenic animals, both transgenes were expressed in independent OSNs. This suggests that the mechanism ensuring monogenic expression of OR genes is able to regulate exogenous DNA sequences carrying OR genes. Note, however, that the YAC used included extensive flanking DNA sequences that might harbour regulatory elements involved in this process[71].

Another important feature of OR expression is that the chosen gene is expressed in a monoallelic fashion. An early study used crosses of divergent mouse strains that allowed the identification of the maternal from the paternal allele of two specific OR genes. Using serial dilutions of OSNs, a statistical argument indicated that when the cells were diluted enough, it was likely to have only one OSN expressing the probed receptor; several of these pools were studied and in most cases only the maternal or paternal allele could be identified. Similar number of cells expressed each allele, suggesting no

parent-of-origin bias. Even though not conclusively proven, this data strongly suggested that OR expression was monoallelic. This was further supported by the observation that in cell lines, OR genes are replicated in an asynchronous manner, a process that is observed only for X-linked genes in female cells and imprinted genes, two classes of monoallelically expressed genes. Again, both the maternal and paternal alleles were identified to be replicated first in equal measures[72].

Since then, the monoallelic character of OR expression has been confirmed numerous times by different methods. For example, by combining DNA- and RNA-fluorescence *in situ* hybridisation (FISH), it was observed that in 90% of the cells expressing a particular OR gene, the DNA probe detects two loci while the RNA-FISH gives only one signal that overlaps with one of the DNA loci; this shows that transcription occurs from one allele only[73]. Perhaps a better proof are experiments with transgenic mice, where each allele is tagged with a different reporter. In these animals, coexpression of both reporters in the same cell was never observed[70].

The monogenic and monoallelic character of OR expression is now largely undisputed. However, a critical analysis of the literature reveals several assumptions that have never been conclusively proven[74]. Despite the large number of OR genes present in the rodent genomes, where most of these studies have been performed, all the observations have been limited to a subset of the receptor repertoire, and thereafter generalised as the rule. To date, there are no studies that have indeed tested all ~1250 mouse ORs to confirm that only one is expressed in each OSN. Furthermore, most evidence has come from double *in situ* hybridisation experiments, or dual tagging with reporter genes; in these cases, only some combinations of receptors have been tested. If some OSNs were to express two (or a few) receptors, and the co-occurrence of any two given ORs was random, the number of OSNs expressing any given combination would presumably be extremely low and, therefore, almost impossible to observe with these methods.

ORs are expressed in zones within the MOE.

Evaluation of expression of different OR genes by *in situ* hybridisation readily revealed a characteristic pattern of expression: each receptor was expressed in a confined zone of the epithelium. Some probes hybridised in regions where signal was never detected for other genes. The study of the expression patterns of a few dozen probes for different subfamilies led to the identification of four broad zones[61], each comprising about a quarter of the surface of the neuroepithelium[66]. The different OR genes within a subfamily were expressed in the same zone and very few labelled cells were found outside this region.

Interestingly, within a zone, the expression of each OR was dispersed and showed no obvious clustering or organisational pattern. The observed expression was symmetrical between the two nasal cavities and was remarkably similar between different individuals regardless of sex[18, 66, 67].

These four zones are organised along the dorso-ventral (DV) and medial-lateral axes of the epithelium. They constitute bands covering different parts of the septum and turbinates, and are continuous along the antero-posterior (AP) axis[66, 67, 75]. Zones were numbered from 1 to 4, with 1 being the most dorsal and 4 the most ventral[61] (Figure 1.6). Paralogous ORs tend to be expressed in the same zone, but there is not a perfect correlation. The ORs expressed in a particular zone map to different clusters throughout the genome and genes of the same cluster can be expressed in different zones[61]. Detailed study of the expression pattern of OR genes has been limited to a small fraction of the complete repertoire. Most of the analysed receptor genes conform to the expression paradigm described above but exceptions have been identified. In both mouse and rat, there is a subfamily of ORs that contain an extended extracellular loop 3, referred to as the OR37-related genes. Interestingly, all these genes are expressed exclusively in constrained regions in endoturbinate II and ectoturbinate 3, instead of being scattered along a whole zone; this region of expression has been termed the *patch*[18, 68]. Class I ORs are mostly found within zone 1, in the most dorsal domain of the epithelium, scattered across the whole zone, and intermingled with some class II ORs[76].

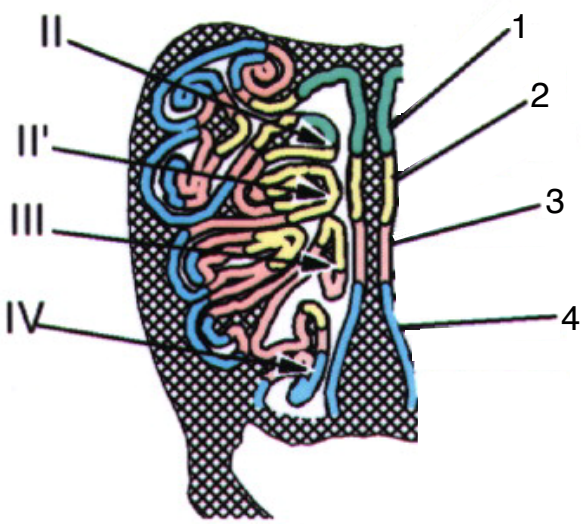


Figure 1.6 – Olfactory receptors are expressed in zones. Schematic of a coronal section of the MOE. The endoturbinates are indicated. The zones 1 to 4 are delineated by different colours. Figure reproduced from [77] by permission of Oxford University Press.

Years after the initial characterisation of the OR expression patterns, a study was performed with probes for 80 different class II OR genes. The expression of these different genes revealed that ORs expressed in zones 2-4 are organised in a continuous and partially overlapping manner, along the dorsomedial and ventrolateral axis of the neuroepithelium. What before were considered as clear boundaries separating mutually exclusive zones actually are occupied by genes expressed with different degrees of overlap[78]. Even though the four broad zones are still held as a conceptual model for the organisation of the expression makeup of the MOE, increasing evidence supports that zone 1 is separated from the rest of the epithelium, which contains many expression bands with varying degrees of overlap. This model is, however, based on signals from less than 10% of the repertoire and could still be incomplete.

OSNs create a topographic map for odorant recognition.

OSN axons coalesce into sites in the MOB called glomeruli, which are spherical congregates of neuropil of varying size; here, OSNs synapse with the MOB's mitral and tufted cells[6, 79]. Every animal contains two bulbs, and each can be divided into two halves, one medial and one lateral; therefore, each individual contains four half-bulbs[80]. There are approximately 1600-1800 glomeruli within each bulb[60]. From the expression patterns of particular OR genes, it is clear that neurones expressing the same OR are dispersed across a region of the MOE. How, then, does the MOB identify which OSNs have been activated upon odorant stimulation? *In situ* hybridisation experiments not only labelled the OSN's soma in the MOE, but also their axons and the glomeruli in the MOB. Probes detecting a small number of OR genes each revealed a few labelled glomeruli[79, 81]. Different probes hybridised with distinct sets of glomeruli that never overlapped, even when the genes were expressed in the same zone in the epithelium[79]. The positions of the labelled glomeruli were bilaterally symmetrical between the two bulbs and were found at roughly the same positions in different animals[79, 81]. However, the glomeruli for a specific receptor can vary in their relative location by a few glomeruli between individuals[70, 82], with enough variation such that it is not possible to determine the identity of a glomerulus just by its location in the MOB[60].

Genetic engineering of mice allowed a precise characterisation of the projection paths from the MOE to the bulb. By coexpressing the fusion protein tau-lacZ –which is transported down axonal processes– from the locus of the *P2* (*Olfr17*) OR gene, it was possible to stain the OSNs expressing this receptor with strong signal throughout the axon. These processes could be observed leaving the MOE through the cribriform plate,

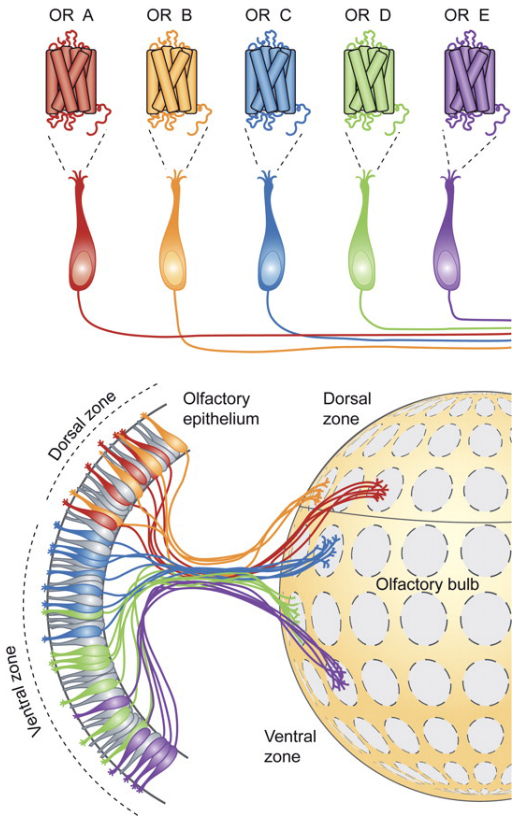


Figure 1.7 – OSN axons coalesce into glomeruli. Schematic representation of five different ORs; each patterns a specific subpopulation of OSNs, as indicated by the corresponding colour. The axons from all the OSNs expressing the same OR coalesce into a particular glomerulus in the main olfactory bulb. Different ORs coalesce into different glomeruli. Reproduced from [83].

entering the outer nerve layer of the MOB and finally coalescing into distinct glomeruli in the glomerular layer of the MOB (Figure 1.7). All visualised axons converged into two glomeruli, one in the medial and one in the lateral halves of each bulb, with no axonal fibres observed anywhere else[84]. Several other OR genes have been engineered in a similar way and support these findings: all the OSNs expressing a particular OR gene are scattered throughout an epithelial zone, and send their axons into two glomeruli per bulb. Different ORs always coalesce into mutually exclusive glomeruli. A few exceptions have been identified, where only one glomerulus per bulb is labelled[70]. Thus, since each OSN most likely expresses a single OR gene, and all the OSNs expressing the same OR synapse at the same glomeruli in the MOB, a topographic map is constructed in the bulb; this allows the identification of which OSNs have been activated and, therefore, the nature of the stimulus. In other words, the task of odour recognition is reduced to identifying which glomeruli have been activated.

A topographic map that links ORs to information processing centres is a good design

to make sense of the diverse stimuli encountered by animals. However, it poses a complex problem of axonal wiring; axons from scattered neurones must find their way into localised points in the MOB, while navigating through axons from more than other thousand different types. Initial experiments with transgenic animals showed that the OR expressed by a given OSN is an important determinant for its axonal projection. OR proteins, though abundant in the cilia of the OSN, are also present in the axons[85]. Several replacement experiments were performed, where the CDS of a given OR was replaced by that of a different one. Very often this resulted in the generation of novel glomeruli, which were different from both the donor and recipient ORs glomeruli; this was independent of whether the donor OR was expressed in the same or a different epithelial zone as the recipient locus[80, 85, 86]. However, there is one example of swaps between two very similar ORs that did not cause formation of novel glomeruli. M71 (*Olfr151*) and M72 (*Olfr160*) are 96% identical; in an animal containing the CDS of *M71* (or *M72*) in the locus of *M72* (or *M71*), the axons of OSNs expressing both the endogenous M71 and the M71→M72 receptors coalesced into the same glomeruli[80]. Interestingly, alterations of the amount of receptor protein do have an impact in axon convergence. A mouse where *M71* was translated from an internal ribosome entry site (IRES) had 68% reduced M71 protein expression compared to control OSNs[87]; in these animals, the OSNs expressing lower amounts of M71 coalesced into glomeruli that were different to those expressing normal levels of M71[85].

Furthermore, the CDS of the *M71* OR was replaced by that of the β2 adrenergic receptor (β2AR), a 7 transmembrane GPCR that shares some of the conserved features observed in ORs and that is able to couple to Gα_{olf}; OSNs that expressed this gene did so in the typical punctate pattern observed for ORs, and their axons converged into specific glomeruli. This demonstrated that the formation of glomeruli does not require an OR able to transduce olfactory information. However, not any GPCR was able to instruct glomerular formation, since this did not happen when the replacement was with a vomeronasal type 1 receptor[85].

Positional cues also play a role in glomerular organisation. Along the DV axis, there is a strong correlation between the positions of the OSNs in the MOE and their corresponding glomeruli in the MOB[78]; therefore, changing the expression domain of an OR in the MOE results in a corresponding shift of its glomeruli in the MOB[78, 88]. The differential projection to each area is achieved by two sets of axon guidance molecules: ROBO2/SLIT1 and NRP2/SEMA3F. Expression of the receptor *Robo2* forms a gradient with highest expression in the dorsal and lowest in the ventral regions of the MOB. In

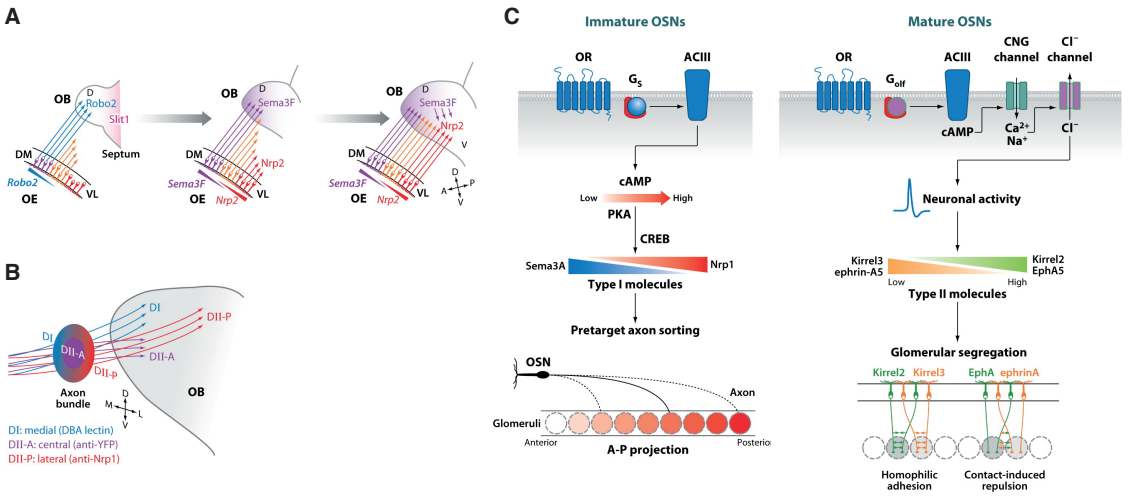


Figure 1.8 – Glomerular organisation in the olfactory bulb. A) A model for the DV projection of OSN axons. The development of the MOB starts from the dorsal domain and extends ventrally. Thus, axons from OSNs from the dorsal MOE reach the bulb first. *(Left)* These express Robo2 which is repelled by Slit1 and thus are confined to the dorsal MOB. *(Middle)* Dorsal OSNs express Sema3F and deposit it on the dorsal MOB. *(Right)* Axons from ventral OSNs express Nrp2; when they reach the MOB, are repelled from the dorsal domain through interaction with Sema3F. **B)** Pretarget axon sorting of OSNs. Axons from OSNs in the dorsal MOE innervate the dorsal MOB. Here, sorting occurs depending on the receptor class, with class I ORs innervating the most dorsal region of the MOB (DI) and class II ORs occupying the zone just ventral to that (DII). Class II ORs are further segregated into an anterior and posterior domains (DII-A and -P). The axons are already sorted in the axon bundle before they reach the MOB. **C)** Activity-dependent axon sorting. *(Left)* Each OR generates a specific level of cAMP; this in turn results in differential levels of Nrp1 and Sema3A, which are expressed in complementary gradients along the AP axis and determine the sorting of glomeruli. *(Right)* Further, different ORs generate different levels of neural activity which determine the level of expression of Kirrel2, Kirrel3, EphA5 and ephrin-A5 in OSNs. These molecules then participate in axon sorting to ensure glomerular segregation of the different OSN types. Figure taken from [25].

contrast, the ROBO2 ligands *Slit1* and *Slit3* are expressed primarily in the ventral part of the bulb, suggesting that the *Robo2*-expressing axons might be targeted to the dorsal bulb through SLIT-ROBO repulsion mechanisms. Consistent with this, in animals that lack expression of either *Robo2* or *Slit1* the dorsal axons are found in more ventral regions of the MOB[89] (Figure 1.8A).

A similar mechanism involving the receptor Neuropilin-2 (*Nrp2*) and its ligand Semaphorin-3F (*Sema3*f) also operates to establish the correct separation of axons along the DV axis. These guidance molecules are expressed in an opposite fashion, with the receptor *Nrp2* highest in the ventral part of the MOE and MOB and the ligand *Sema3f* highest in the dorsal aspect. Expression of *Sema3f* is observed from embryonic day (E)14.5 which precedes arrival of axons to the MOB. The development of the MOB during embryogenesis starts from the dorsal side and extends ventrally. Therefore, *Robo2*-expressing axons from the dorsal epithelium are the first to innervate the developing MOB; once in there, they excrete SEMA3F. The late arriving axons from the ventral region of the MOE, that express NRP2, are then repelled by SEMA3F and are therefore confined to the ventral domain (Figure 1.8A). Consistent with all these, knockout of

Sema3f results in the mistargeting of *Nrp2*$^+$ axons to the dorsal domain of the MOB. Similarly, overexpression or knockout of *Nrp2*, shifts the glomeruli ventrally or dorsally, respectively[90].

There is a further subdivision in the dorsal domain of the MOB, depending on whether the OSNs express class I or class II ORs. The OSNs that express class I receptors project to the dorsal-medial aspect of the MOB (termed DI), while OSNs expressing class II ORs are found in the dorsal-lateral region (DII). These two classes of OSNs are intermingled throughout the dorsal MOE but as their axons exit towards the MOB there is a segregation in the axon bundle depending on the class of the OSN, before reaching the MOB (Figure 1.8B). Interestingly, the class of the OSN is not defined by the OR protein, but by the locus of expression; that is, in swap experiments where the coding sequence of a class I gene is inserted in the locus of a class II OR, the projections are to the class II domain[91].

In the AP axis of the bulb there is no correlation with the MOE, since OSNs expressing particular ORs are scattered along this axis without any evident organisation. Nonetheless, there is a clear segregation of particular OR species into distinct regions in the MOB. The study of a mouse strain that contains a mutated *I7* (*Olfr2*) OR gene incapable of coupling to Gα_{olf} revealed that the axons of OSNs expressing such receptor failed to reach the glomerular layer of the MOB and form glomeruli; this suggested a role for the production of cAMP in glomerular formation. Indeed, depending on the levels of cAMP produced, the glomeruli were positioned differentially. A gradient is apparent with cAMP levels high in the posterior and low in the anterior part of the MOB[92]. Different ORs have different spontaneous firing rates when devoid of odorants[93] and also different levels of *Adcy3* expression[94] which altogether result in varying levels of cAMP. Furthermore, several guidance molecules are differentially expressed depending on the cAMP levels produced by the OSN, such as Neuropilin1 (*Nrp1*) which is also found in a posterior-high anterior-low fashion[92]. Interestingly, the graduated expression of *Nrp1* is evident already in the axon bundle, before the MOB is reached. *Nrp1* is the receptor for the repulsive ligand Semaphorin-3A (*Sema3*a) and, correspondingly, they are expressed in a complementary manner along the AP axis. The interaction between these two molecules separates the anterior from the posterior domains (Figure 1.8C) and alteration of the levels of either molecule results in a disorganisation of the glomeruli into ectopic locations[95]. Therefore, the graduated expression of signalling cues allows a crude arrangement of OSN axons expressing different receptors to coalesce into distinct regions, based on their position in the MOE (DV axis) and their levels of

cAMP (AP axis).

Further refinement and pruning occurs after birth, in an activity-dependent manner. It is common to transiently observe multiple glomeruli for the same OR gene during development; these aren't homogeneous and axons from OSNs expressing other ORs are found within. The rate at which such glomeruli are refined into a single, homogenous structure varies for different ORs. If the sensory stimulation is prevented by surgically closing one of the nostrils (a procedure referred to as unilateral naris closure or occlusion) the refinement doesn't occur in the deprived side and multiple glomeruli are still present in adults[96]. Another set of molecules that are expressed in an activity dependent manner are KIRREL2, KIRREL3, EPHA5 and ephrin-A5 (*Efna5*). EPHA5 and ephrinA5 have been shown to interact with each other and provoke repulsion; consistently, they are expressed in a mutually exclusive manner. *Kirrel2* and *Kirrel3* are also found expressed in complementary sets of OSNs: when one is high, the other is low (Figure 1.8C). The expression of these genes is dependent on neuronal activity. In a *Cnga2* knockout animal, lack of expression of this channel correlates with high expression of *Kirrel3* and *Efna5* and no expression of *Kirrel2* and *Epha5*; similar results were obtained by naris occlusion. In contrast to the molecules described above, these genes are not expressed in a gradient across the MOE but, instead, show a mosaic pattern determined by the OR gene expressed. A swap of the coding sequence of one OR into the locus of a different one also alters the levels of *Kirrel2* and *Epha5*. Based on these data, a model has been proposed whereby the initial sorting of axons in the AP axis is guided by NRP1 and SEMA3A; further refinement is achieved by repulsion of axons from OSNs expressing different ORs by the distinct expression of EPHA5 and ephrin-A5, and attraction of axons expressing KIRREL2 or KIRREL3[97] (Figure 1.8C). Other molecules yet to be identified might also be involved in these processes.

Trace-amine associated receptors.

Screening of an OSN cDNA library with probes for other GPCRs, not previously identified as chemoreceptors, revealed that genes from the trace amine-associated receptor (TAAR) multi-gene family were present in OSNs. By *in situ* hybridisation experiments, the expression was observed in a subset of OSNs of the MOE, scattered in certain domains of the epithelium; a pattern that resembled that of OR genes[98]. Furthermore, the expression was abundant in the dendrites, supporting their role in chemosensation[99]. OSNs that expressed *Taar* genes also expressed all the components of the canonical signalling pathway (*Adcy3*, *Gnal*, *Cnga2, Ano2*) which implies that these OSNs use the

same transduction mechanism as OR-expressing OSNs[100].

Double *in situ* hybridisation with probes for different *Taar* genes revealed that each probe labeled a distinct subpopulation of OSNs, suggesting these genes are also expressed in a monogenic fashion. Consistent with this, no evidence of coexpression with several OR genes was found, though only a few genes were tested[98, 99]. Additionally, when both alleles of *Taar4* were tagged with two different fluorescent reporter proteins, no coexpression could be observed, suggesting monoallelic expression [101].

The mouse genome contains 15 *Taar* genes, all located in a single cluster in chromosome 10; 14 are expressed in the OSNs of the MOE. This class of GPCRs is not related to ORs and their closest relatives are receptors for biogenic amines such as serotonin and dopamine[98, 102]. Ten out of the 14 *Taar* genes expressed in the MOE were found in the dorsal part of the epithelium, intermingled with the class I and class II ORs found there; two more were located ventrally and the remaining were in both zones. All the dorsal *Taars* send their axons to several specific glomeruli in the dorsal MOB, in between the glomeruli from the class I and class II dorsal ORs [99, 101]. Experiments showed that when an OSN chose a non-functional *Taar* gene –for example because the coding sequence was substituted by a LacZ cassette– a second receptor was chosen. However, these cells were strongly biased towards choosing another *Taar* gene, and very rarely chose an OR. Moreover, the allele chosen was preferentially selected from the other chromosome, which suggests that the bias was not due to a positional bias where nearby genes were more likely to be chosen. [99, 101]. Interestingly, the number of neurones that express *Taar* genes is somehow coded in the choice process, since when the *Taar* cluster was deleted from one chromosome, the same number of neurones expressing *Taar* genes was observed compared to wild-type animals[101].

As their name indicates, TAARs are able to bind trace quantities of amines. Expression of several genes in heterologous systems revealed that TAAR4 responds to β-phenylethylamine, which is found in the urine of several species. In mice it increases in response to stress[98, 103] and it is much more abundant in the urine of carnivorous species. Stimulation with this compound activated several glomeruli in the MOB, the number of which increased with increasing concentration; this suggests that there are several receptors responding to it, with differing sensitivities. Several of the activated glomeruli were innervated by *Taar4*-expressing OSNs[103]. *In vivo* recordings from these cells revealed that they were incredibly sensitive and could be activated with subpicomolar concentrations of β-phenylethylamine[104]. Mice are naturally repelled by predator urine; the same behaviour was observed when β-phenylethylamine alone was

used as a stimulus. Supporting the sufficient and indispensable role of this compound in avoidance behaviour, predator urine depleted of β-phenylethylamine no longer repelled mice[103]. What's more, the behaviour was also lost in a *Taar4* KO mouse line[100].

Taar5 is activated by trimethylamine, a compound that is present much more abundantly in mouse urine of adult males compared to females[98, 105]. Mice are attracted to trimethylamine when present a the relevant physiological concentrations. Such attraction was lost in mice lacking the *Taar5* receptor, and the same occurred if the urine was depleted of trimethylamine[105]. A compound similarly found in mouse urine in a sexually dimorphic manner, isoamylamine, was shown to activate *Taar3*[98]. Therefore, data so far indicates that many, if not all *Taar* genes are activated by amines, though some are able to respond to other chemical classes with low sensitivity. Interestingly, *in vivo* recordings from neurones expressing two different genes, *Taar3* and *Taar5*, revealed that both receptors are broadly tuned and can respond to several, structurally diverse amines, albeit at high concentrations; if the concentration is decreased, they become specific to their high-affinity ligand(s)[104]. Stimulation with amines resulted in activation of the *Taar*-innervated glomeruli in the MOB, a response that was lost if the *Taar* genes were deleted, which again suggests the these genes are the primary detectors of amines[100, 101].

Guanylyl cyclase D.

Guanylyl cyclases (GC) are receptors that can be either soluble or membrane bound. The latter contain a single membrane-spanning domain, an extracellular ligand-binding domain and an intracellular region that has a protein kinase-like and a cyclase catalytic domains. Identification of a couple of these receptors in the eye prompted their study in the olfactory system. PCR with degenerate primers identified a novel member of the gene family, named GC-D. In the intracellular region, it showed 40-45% identity with other known GCs but the extracellular domain was very different (16-21% identity). Its expression was assessed by *in situ* hybridisation; individual OSNs were labelled in the central region of the four turbinates of the MOE, in a similar manner as is observed for ORs[106]. The localisation of the protein was mainly to the olfactory cilia, consistent with a role in chemosensation[107]. RT-PCR and Northern blot hybridisations revealed GC-D was specifically expressed in the MOE and could not be detected in cDNA from other tissues[106].

GCs can bind peptides through their extracellular domain and this leads to the production of cGMP. This raised the possibility that GC-D-expressing OSNs might be using

a cGMP transduction pathway for olfactory signalling. Cyclic nucleotide phosphodiesterases (PDEs) are able to hydrolyse second messengers such as cAMP and cGMP. Immunohistochemistry and *in situ* hybridisations of the MOE revealed a subset of OSNs labeled with probes for PDE2, which were also positive for GC-D. This suggests that activation of GC-D, which leads to an increase in cGMP levels, could stimulate PDE2. These cells were negative for the canonical –cAMP mediated– signalling proteins, such as *Adcy3*. PDE2 was expressed in the cilia of the OSNs along with GC-D, but it was also present in the axons. Therefore, labelling neurones with this gene revealed the axon bundles projecting to a group of glomeruli in the caudal region of the MOB; these are termed the necklace glomeruli because they are interconnected by nerve fibres and resemble a beaded necklace[107]. If the GC-D-expressing OSNs were indeed using a cGMP based signalling pathway, a cGMP-selective CNG channel should be expressed in these cells. A previously identified subunit of a CNG channel that is cGMP-selective was found to be expressed in a subset of OSNs in the MOE (*Cnga3* in mice), preferentially in their cilia; these labelled cells were confirmed to express also GC-D and PDE2 and lack the canonical signalling proteins[108].

Further to the signalling components of the cGMP-based pathway, GC-D$^+$ neurones also express high levels of carbonic anhydrase type II (CAII), a gene not found in other OSNs. Consistently, its expression is also observed in the necklace glomeruli. CAII catalyses the conversion of carbon dioxide (CO_2) and water into bicarbonate and protons. Responses to CO_2 were confirmed both in GC-D$^+$ OSNs and in the necklace glomeruli, in a dose-dependent manner. Activation was only observed in the presence of extracellular calcium and intact CNGA3 and CAII. Further behavioural tests demonstrated that mice were able to detect CO_2, at near atmospheric levels, and learnt to associate it with a reward[109]. Expression of GC-D in a heterologous system revealed that the intake of bicarbonate resulted in an increase of cGMP, through the cyclic catalytic domain of GC-D. Therefore, the stimulation of GC-D$^+$ OSNs with CO_2 results in the production of bicarbonate, through CAII; this in turns activates GC-D which produces cGMP; an increase in cGMP then opens the CNG channel CNGA3 to allow an influx of cations into the neurone and elicit an action potential[110].

In a similar manner, GC-D$^+$ neurones were shown to be able to respond to carbon disulfide (CS_2), which is found in mice breath, and can also be processed by CAII. Concentrations of CS_2 in the sub-micromolar range were enough to elicit a response, indicating that GC-D$^+$ OSNs are much more sensitive to this chemical than to CO_2. Mice learn which foods are safe to eat by smelling the breath of conspecifics, a process

known as social transmission of food preference (STFP)[111]. This phenomenon requires the presence of CS_2 paired with food odours, both found in the breath of an animal that has recently ingested food. Interestingly, animals lacking either GC-D or CNGA3 were unable to show learned food preference, directly implicating GC-D$^+$ neurones ability to respond to CS_2 in this behaviour[112].

Further studies identified that the GC-D$^+$ neurones are activated upon stimulation with uroguanylin, a peptide present in mouse urine, and the related peptide guanylin. These responses were directly dependent on GC-D expression and the presence of a functional CNG channel. Interestingly, different subpopulations of GC-D$^+$ neurones could be identified; around half of them were activated by both peptides, and an additional quarter were specifically responsive to one but not the other[113]. It has also been observed that mice prefer food sources that are in close proximity to conspecific social odours. Mice show a strong preference to feed in places where other mice have deposited urine and faeces, which is a sign that the food is safe to eat. Uroguanylin is excreted in both urine and faeces and its concentration increases upon feeding. Thus it has been proposed that its recognition by GC-D$^+$ neurones could also be related to food preference. Indeed, when odourised food was presented along with uroguanylin, mice showed a strong preference for that particular odour, in a similar manner as they would if presented with the faecal pellets of mice that consumed the odourised food. This behaviour was dependent on GC-D[114].

1.1.2 The vomeronasal organ.

The vomeronasal organ (VNO), also known as Jacobson's organ, is a paired tubular structure confined within a bony capsule. It is located at the base of the nasal septum (Figure 1.1), which divides it into symmetrical halves, each containing a crescent shaped lumen surrounded by cavernous tissue. It is connected to the nasal cavity and in some species there is also an opening to the oral cavity. The air flow from respiration does not contact the VNO directly; it is instead stimulated by non volatile molecules that require direct contact with the animal's snout for detection. Next to the lumen there are blood vessels that through vasodilation and vasoconstriction generate a pumping action that helps transport the stimuli into the lumen, which is filled with fluid. Within the cavernous tissue can be found many glands with secretory ducts that end in the lumen. The concave side of the lumen is lined by a pseudostratified neuroepithelium that contains, similar to the MOE, sensory neurones, sustentacular and basal cells (Figure 1.9).

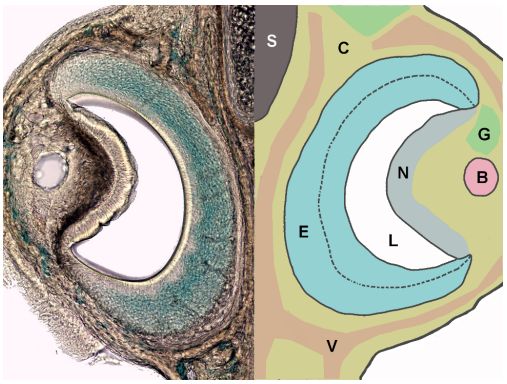


Figure 1.9 – The mouse vomeronasal organ. A coronal section through half of the VNO of an adult mouse *(left)* with a cartoon of the corresponding tissue morphology *(right)*. S, nasal septum; C, cavernous tissue; G, glandular tissue; B, blood vessel; V, vomer; N, nonsensory epithelium; L, lumen; E, sensory epithelium with apical *(right)* and basal *(left)* layers of vomeronasal sensory neurones. Figure reproduced from [1] with kind permission from Springer Science and Business Media.

Vomeronasal sensory neurones (VSNs) are bipolar cells that extend a dendrite to the surface of the epithelium. Such dendrite terminates in a vesicular structure that is covered with microvilli; analogous to the OSN cilia, these structures are the interaction point with chemicals. From the opposite pole, a single axon travels through the cribriform plate into the accessory olfactory bulb (AOB), which is located in the posterior dorsal part of the MOB (Figure 1.1). Basal stem cells are found towards the boundary with nonsensory epithelium and have the capacity to proliferate and differentiate into VSNs throughout the animal's life[11, 115, 116].

Vomeronasal signalling.

Semiochemicals that reach the VSNs need to be recognised and their identity must be transmitted to the AOB. Three families of receptor genes have been identified in the mouse VNO —two families of vomeronasal receptors (V1Rs and V2Rs) and a group of formyl peptide receptors (Fprs)– and some evidence exists to support their role in binding olfactory cues. Communication between the VNO and the AOB is initiated by the receptors binding their cognate ligand; this triggers a signal transduction pathway that results in the generation of an action potential in the stimulated VSNs. Initial efforts to characterise the signalling cascade focused on the genes involved in the same process in the MOE; none of these could be detected in the VNO. A search for analogous components led to the identification of the G-protein α subunits $G\alpha_{i2}$ and $G\alpha_o$ (Figure 1.10). These are highly expressed in VNO neurones, in two mutually exclusive populations; VSNs that express $G\alpha_{i2}$ are located in the apical region of the neuroepithelium while the ones expressing $G\alpha_o$ sit in the basal portion [117]. For both cellular populations,

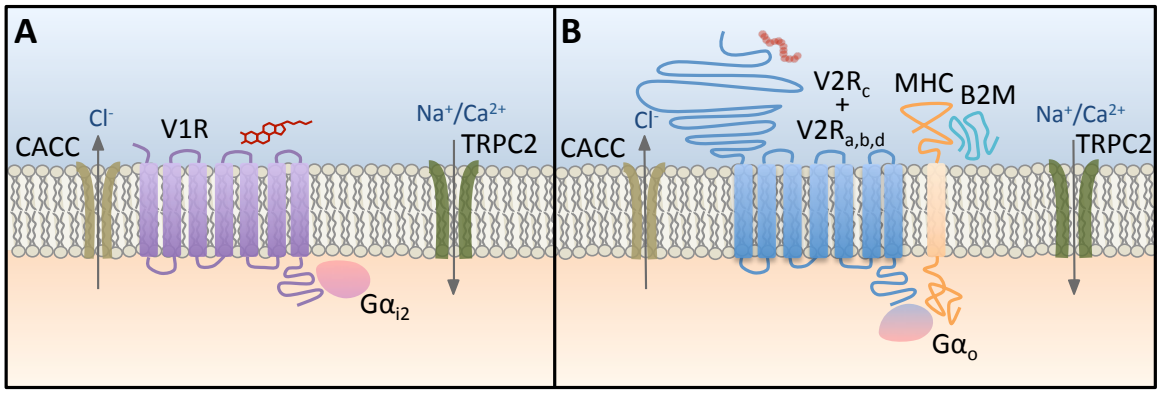


Figure 1.10 – Signal transduction proteins in vomeronasal sensory neurons. There are two subclasses of mammalian vomeronasal sensory neurones (VSNs). **A)** In apical VSNs, a V1R receptor associated with the $G\alpha_{i2}$ G-protein subunit is activated by a small, volatile chemical ligand. **B)** In basal VSNs, a V2R receptor from subfamily C is coexpressed with one from subfamily A, B, or D. These are associated with the $G\alpha_o$ G-protein subunit and are individually or collectively activated by a peptide or protein ligand. One or more of nine major histocompatibility complex (MHC) class 1b proteins and β2-microglobulin (B2M) are also expressed in a subset of these neurones. Both types of neurone additionally express a transient receptor potential ion channel (TRPC2) and calcium-activated chloride channels (CACCs), which together depolarise the cell. Figure reproduced from [1] with kind permission from Springer Science and Business Media.

expression is localised to the microvilli of the neurones, where ligand detection occurs.

The functional importance of both subunits in mediating behavioural responses was established by ablating the genes in mice. $G\alpha_{i2}$-mutant males displayed a diminished aggressive response in a classical 'resident-intruder test', when an intruder male was introduced to the cage of a territorial resident. Likewise, mutant lactating females were also less aggressive, but sexual behaviours appeared unaltered [118]. However $G\alpha_{i2}$ is expressed in other tissues and the mutant animals had other debilitating phenotypes [119]; therefore it remains possible that the aberrant behaviour observed was not a direct consequence of VNO-mediated signalling. With this caveat in mind, Chamero et al.[120] generated a mutant line with $G\alpha_o$ ablated only in vomeronasal neurones. These animals displayed strikingly similar behaviour to $G\alpha_{i2}$ deficient mice in that both sexes were less aggressive[120]. Thus both classes of VSN appear to transduce chemosensory-mediated aggressive behaviour.

In 1999, Liman et al.[121] identified another key player in eliciting VNO signal transduction: a member of the transient receptor potential (TRP) family of ion channels, *Trpc2*. The rat *Trpc2* gene was shown to be abundantly expressed in the VNO. Detailed analysis demonstrated that the protein was found in the microvilli of the sensory neurones, and colocalised with expression of both $G\alpha_{i2}$ and $G\alpha_o$ [122]. The dramatic role of *Trpc2* in vomeronasal-mediated behaviour was made evident when the gene was knocked out in mice. Two groups independently showed that VSNs from these animals

were either non-responsive, or had a significantly reduced response to urinary semiochemicals. Behavioural analyses of the mutant males revealed a diminished aggressive response in the resident-intruder paradigm. Instead of initiating an attack, *Trpc2*$^{-/-}$ males displayed sexual behaviour towards the intruder, just as a *Trpc2*$^{+/+}$ mouse does when presented with a female. Additionally, when presented with both a male and a female, *Trpc2*$^{-/-}$ males did not discriminate between them[123, 124]. These led to the conclusion that these mice are unable to determine the sex of the conspecifics they encounter due to the lack of signal transduction of olfactory cues through VSNs.

However, residual electrophysiological activity could still be detected in the VNO of *Trpc2*$^{-/-}$ animals, suggesting additional ion channels are present in VSNs; these were later identified as calcium-activated chloride channels (CACCs) [125]. Consistent with this, elimination of intracellular Cl- reduced the response of VSNs to urine stimuli and completely abolished residual urine-evoked currents in *Trpc2*$^{-/-}$ neurones. Although activity of these channels are both necessary and sufficient for activation of the VSNs[126], it is *Trpc2*$^{-/-}$ mice that have proven most useful for revealing additional VNO-mediated behaviours. Like males, *Trpc2* mutant lactating females are not aggressive toward intruder males and are deficient in maternal behaviours[123, 127]. Also, they display male-like sexual behaviours towards intruders, such as mounting and pelvic thrusts [128]; as with male residents, when mutant females are presented with both male and female intruders, they show no preference towards mounting one sex. Thus *Trpc2* appears necessary for VSNs to effectively transduce a range of chemosensory cues that are transmitted between mice to initiate social behaviours. More recently *Trpc2*$^{-/-}$ mice were used to demonstrate that VSNs also detect olfactory cues from other species [129]. The mutant mice do not display innate defensive and avoidance behaviours, or a stress response, when exposed to predator cues from snakes, cats and rats [130].

A caveat of all these studies is that, historically, *Trpc2* has been considered to be specifically expressed in the VNO and virtually absent in the MOE. While this is true in rats [121], mice have a different expression pattern. It has recently been shown that a population of neurones in the MOE express *Trpc2* from embryonic day E16.5 throughout adulthood. It was further demonstrated that the positive cells contain the protein product and that at least some of the neurones' axons coalesce into a few glomeruli in the ventral region of the MOB, near the necklace glomeruli. This suggests that the positive cells are indeed neurones[131]. These findings, therefore, question the interpretation of results obtained through Trpc2$^{-/-}$animals, since it can no longer be assumed that all the behavioural dysfunctions observed are due to VNO mediated signalling.

The vomeronasal receptor genes.

The vomeronasal receptor (VR) genes are encoded by two multigene families of GPCRs. These are not closely related to the ORs expressed in the MOE and, furthermore, are independent of each other in their evolutionary origins. Under the assumption that the receptors of the VNO might be expressed in a similarly monogenic fashion as observed for ORs in the MOE, Dulac and Axel devised a clever differential hybridisation strategy that allowed them to find coding sequences expressed specifically in one VSN, but not others. With this methodology they recovered the coding sequence for a gene encoding a putative seven transmembrane domain, that was expressed in a subpopulation of VNO neurones [132]. Additional related genes were then identified and it was confirmed that they were part of a multigene family. Each of the receptors tested by *in situ* hybridisation was expressed in a subset of neurones, similar to the expression patter of ORs in the MOE. Interestingly, expression could only be detected in the apical, $G\alpha_{i2}{}^{+}$ region of the neuroepithelium. All the above suggested that these genes were putative receptors, and that each VSN likely expressed a single receptor gene [132]. This receptor family comprises the V1Rs. A couple of years later, three different groups reported the expression of an unrelated multigene family of receptors expressed in the basal, $G\alpha_{o}{}^{+}$ portion of the VNO. These were similarly expressed in a small subpopulation of VSNs suggesting also monogenic expression [133–135]. These receptors were termed V2Rs.

V1Rs.

With the availability of a good mouse reference genome, it has been possible to identify the complete receptor repertoire. The mouse genome contains 392 V1R genes (named *Vmn1r* in mice), 239 of which have an intact open reading frame (ORF) [136]. A phylogenetic tree constructed with 137 of the intact genes, groups them into twelve distinct subfamilies (*Vmn1ra-j*; Figure 1.11). Receptors from the same subfamily share at least 40% identity at the amino acid level, but the diversity between different families is large, and identities can be as low as 15% [137]. *Vmn1r* genes of the same subfamily tend to be found together in the genome arranged in tight clusters of genes; these are then dispersed across several chromosomes [138].

V1Rs have been shown to respond to low-molecular-weight organic molecules with great sensitivity. Screening of VSNs with six different chemicals with putative pheromonal activity showed that each activated a small subset of neurones [139], and at least one of them was able to generate responses in neurones expressing different *Vmn1rs*

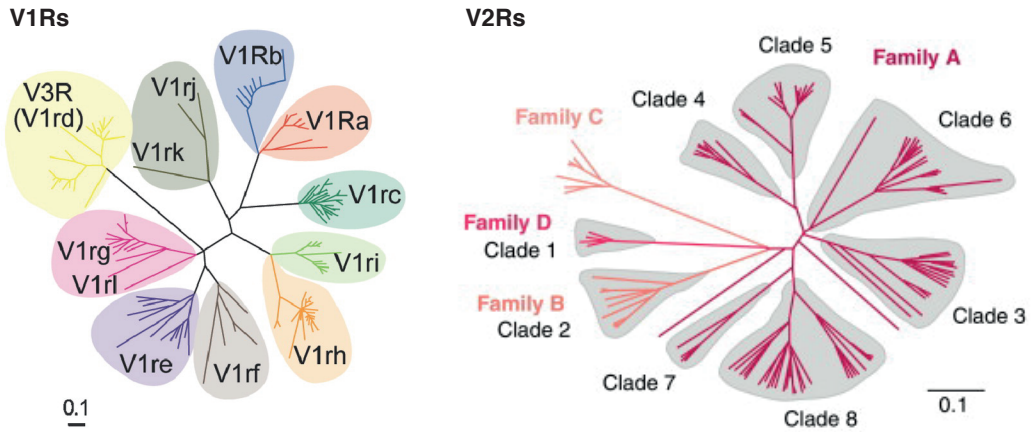


Figure 1.11 – The mouse vomeronasal receptor gene family. Phylogenetic trees of the V1R *(left)* and V2R *(right)* gene families. V1R tree reproduced from [144]. V2R tree reprinted from [145], copyright (2012) with permission from Elsevier.

[140]. Exposure of VSNs to sulphated steroids, which are present in female urine and are proposed to account for most of its vomeronasal bioactivity, resulted in the firing of both male and female *Vmn1r*-expressing VSNs. While some receptors responded to specific steroids, others recognised several compounds that were chemically related [141, 142]. To characterise the behavioural role of *Vmn1r*-expressing VSNs, a group of 16 intact receptor genes belonging to the families *Vmn1ra* and *Vmn1rb* were deleted in the mouse genome by chromosome engineering. Mutant female animals showed deficits in maternal aggression towards intruders and mutant males had lower mating rates [143]. Therefore, at least some of these receptors are necessary for the normal display of innate behaviour.

V2Rs.

The mouse reference genome contains 279 V2R genes (termed *Vmn2r* in mice), 158 of which are characterised as pseudogenised [65]. The predicted intact sequences can be grouped into four different subfamilies (A-D). Most of the genes (85%) belong to the A subfamily, which is further subdivided into nine clades (Figure 1.11). As with *Vmn1rs*, genes closely related tend to be clustered in the genome[146]. *Vmn2rs*, however, are distinct in their expression logic. Each VSN of the basal VNO expresses a member of the subfamily C (composed by seven genes in the mouse), along with an additional *Vmn2r* gene from subfamily A, B or D in a non-random fashion [147–149]. In addition to this, basal VSNs have been shown to express genes of the major histocompatibility complex (MHC) class 1b and β2-microglobulin (B2m, which is essential for the proper expression of MHC class Ib molecules at the cell surface). These proteins localise to the

dendritic tips of VSNs, as do TRPC2 and Gα_o. Each of the nine genes in this family (M1, M9, M11 and six members of the M10 family) is expressed in a subset of neurones positive for Gα_o; even though most of the neurones express a single gene, some can express two or three. The expression of specific members of this family is linked to pairs of *Vmn2rs* in a tripartite fashion and, along with B2m, they have been proposed to form a protein complex necessary for the transport of the receptor to the plasma membrane [150, 151].

Vmn2rs have been found to respond to water-soluble peptides and proteins that can be found in urine and other bodily secretions of conspecific mice, as well as from other species. The first evidence for this came from the finding that peptide ligands of the MHC class I molecules activate around 1% of the VSNs, all situated in the basal neuroepithelium. The presentation of different peptides leads to activation of different neural populations, which overlap to some extent. It's been shown, for example, that those VSNs that express *Vmn2r26* (also known as *V2R1b*) recognise some of these peptides, but neurones expressing other receptors are also responsive to the same stimuli. The different peptides that activate the same neurones share key residues at anchor positions, and these are necessary and sufficient to induce the response[152, 153]. These peptide cues also induce the Bruce effect in female mice (a selective chemical cue induced pregnancy failure[154]) when spiked into otherwise familiar male urine [152], thus establishing them as a 'signature mixture' of odours [155]. Subsequently, further protein ligands that activate *Vmn2r*-expressing neurones have been identified. These include products of the *Mup* and *Esp* gene families that either encode identity or initiate sexual, attractive, aggressive and avoidance behaviours [130, 156–159].

Formyl peptide receptors.

In order to determine if additional chemosensory receptors were expressed in the VNO, two groups independently prepared cDNA from mouse VSNs and amplified GPCRs that hadn't previously been implicated in chemodetection [160, 161]. Five of the seven members of the formyl peptide receptor (*Fpr*) family were recovered. *In situ* hybridization revealed that each receptor was expressed in a subset of VSNs, in a similar manner to what is observed with *Vmn1rs*. Similarly, no single neurone was patterned by two different *Fpr* genes. The VSNs that expressed four of the five *Fprs* were also positive for Gα_{i2} while expression of a single receptor (*Fpr-rs1*) was restricted to Gα_o positive neurones [160]. No co-expression of *Vmnrs* and *Fprs* could be detected. All these suggest that the VNO contains a third population of VSNs that express a different type of receptor

genes.

N-formylated peptides are found in prokaryotes and mitochondria; accordingly, the other *Fpr* genes are expressed in the immune system and play a role in the host response. Thus it has been proposed that the VNO-expressed *Fprs* may be pathogen chemosensors that elicit avoidance behaviours to resist infection. While this has yet to be demonstrated behaviourally, a number of studies have identified FPR ligands by calcium imaging of VSNs. These include bacterial N-formylmethionine-leucine-phenylalanine (fMLF), the antimicrobial CRAMP and the mitochondrially encoded peptides NDI-6T and NDI-6I[120, 161]. More recently *Fpr-rs1* was found to display stereo-selection for peptides with a D-amino acid in the C-terminal position, further supporting a role in detecting pathogenic chemosignals [162]. *Fprs* are also expressed in the VNOs of rats and gerbils[161], but it is possible that the expansion of the *Fpr* gene family to encompass an olfactory function is rodent specific, as in the genome of primates only the genes expressed in the immune system are found [163].

1.1.3 The septal organ.

The septal organ (SO), also known as the organ of Masera, is a patch of olfactory sensory epithelium located near the ventral end of the nasal septum at the entrance of the nasopharynx[164] (Figure 1.1). It is surrounded by respiratory epithelium that separates it from the caudal end of the VNO and the rostral margin of the MOE[165]. The epithelium of the SO has a similar structure to that of the MOE; it is also a pseudostratified epithelium composed of sensory neurones, basal and sustentacular cells, sitting on top of lamina propria with Bowman's glands[164, 165]. However, the neuronal layer occupied by immature and mature neurones is thinner than in the MOE, with only one to two layers of each type in the SO. Also, the sensory neurones have a flattened somata and shorter dendrites compared to what is observed in the MOE[164].

PCR analysis of SO derived cDNA libraries with degenerate primers to amplify different classes of receptor genes failed to identify expression of any V1Rs, V2Rs or class I ORs; but 120 different class II OR genes could be detected[166–168]. Consistent with this, the neurones expressed *Adcy3* and *Gnal* uniformly across the whole SO, suggesting they are OSNs that use the cAMP signalling pathway, coupled to ORs, to transduce olfactory information[164]. Interestingly, the great majority of these 120 OR genes are expressed in very few OSNs; 11 genes alone account for 95% of the total number of neurones and a single OR is expressed in half of the OSN population[167]. In double *in situ* hybridisation experiments with combinations of these abundantly expressed re-

ceptors, very few or no cells were co-labelled, suggesting that ORs are expressed in a monogenic fashion. All of these receptors are also expressed in the MOE, mainly in the most ventrolateral zone[166, 167]. The SO OSNs project a single axon to the MOB and coalesce into glomeruli in the ventromedial aspect of the posterior bulb[164]. Most fibres coalesce into a few glomeruli, exclusively innervated by axons stemming from the SO, while the rest form a complex network that enters multiple glomeruli mainly composed of axons from the MOE[169].

1.1.4 The Grueneberg ganglion.

A fourth olfactory structure is situated at the rostral end of the nasal cavity, just inside the nostrils, termed the Gruenberg ganglion (GG), since it was initially described as such by Grüneberg in 1973[170] (Figure 1.1). It is composed of only a few hundred round cell bodies, positive for *Omp[170–172]* and *βIII-tubulin* (a neuronal marker)[172]. These cells are clustered in an arrowhead shape, on both sides of the nasal septum[171], under a keratinised epithelium that separates them from the nasal cavity[173]. Additional to the expression of neuronal markers, these cells contain axons that project to the MOB, suggesting they are indeed neurones[170–172]. The structure of the GG is not of a pseudostratified epithelium; instead, cells are tightly packed into clusters without basal or sustentacular cells[171]. However, they are found intermingled with glial cells. The neurones contain cilia but these structures do not protrude into the airspace of the nasal cavity. However, the keratinised epithelium is permeable to hydrophilic molecules, which suggests that water-soluble stimulants might be able to reach the GG cell clusters[173]. The GG becomes apparent from E15.5, with an increase in cell number until E18.5 and appears to be fully developed by birth. It persists throughout adulthood[171, 172]. GG axons form several nerve bundles that travel along the dorsal aspect of the nasal septum into the MOB. Axons innervate several caudal glomeruli that surround the anterior part of the AOB, [170, 171], in the same region where the necklace glomeruli are found[174].

GG neurones, similar to GC-D OSNs, use a cGMP mediated signalling cascade to transduce information; they express receptor guanylyl cyclase G and A (GC-G and GC-A), the phosphodiesterase PDE2A which is stimulated by cGMP, a cGMP-dependent kinase (cGKII) and a cGMP-activated channel (CNGA3). GC-G is expressed in most GG neurones in both the neonate and adult, while GC-A is present only in a small subset of cells scattered throughout the organ[174]. However, they also coexpress $G\alpha_{i2}$ and $G\alpha_o$ and a high proportion express *Vmn2r83*, a V2R gene of subfamily C. No other genes from the V1R or V2R families of receptors have been identified in the GG[175].

Several members of the TAARs are present in a subpopulation of GG neurones, with variable frequencies. The expression of both *Vmn2r83* and the different *Taar* receptors is dynamic with age; the highest numbers of neurones expressing a particular gene are found in prenatal stages with a significant decline into adulthood. Each neurone seems to express only one of these receptor genes[176].

Given that the GG is most prominent in neonates and that its neurones innervate sites close to the necklace glomeruli, it was hypothesised that this structure might be involved in suckling. However, no responses were observed in calcium imaging experiments performed with milk or mammary fluid from lactating females. In contrast, strong responses were recorded when the neurones were stimulated with alarm pheromones (APs), obtained during culling mice with CO_2, which induces stress and the release of these molecules. APs were able to activate GG neurones of both newborn and adult mice. Furthermore, the presentation of APs induces a freezing response in mice and this behaviour was lost when the GG axon bundles were sectioned[173].

Interestingly, additional experiments revealed that most GG neurones also respond to cold temperatures; the calcium increase observed was directly correlated with the decrease in temperature and responses were not observed with exposure to heat[177]. The thermal response of these cells was elicited by activation of GC-G; coolness enhanced dimerisation/oligomerisation of the receptor and this triggered the signalling transduction pathway. In a KO mouse for GC-G, GG neurones were not responsive to coolness anymore. Pups generate ultrasound vocalisations (USV) in response to cool temperature to attract attention from their mother. In GC-G KO pups exposed to coolness, the number of USV calls was significantly decreased and the latency to the first call was substantially increased, suggesting a possible role of the GG thermosensation capabilities in this behaviour[178].

1.2 Regulation of OR expression.

Shortly after the discovery of OR genes it was evident that this multigene family is under tight regulatory control to achieve singular expression in each OSN. Several hypothesis emerged to explain this, involving processes that operate in other multigene families with similar expression patterns. One such proposed mechanism involved gene conversion to translocate a specific OR gene into an active locus[179]. However, the dual DNA and RNA-FISH experiments argued against this, since the DNA probe recognised only two loci, one of which coincided with the RNA probe. If gene conversion

was allowing the expression of the gene, a third location in the genome should contain a DNA signal[73]. A second popular hypothesis was the use of DNA recombination, in a process analogous to the rearrangements observed in the immune system to generate specific immunoglobins and antigen receptors. In this model, recombination events would bring together a promoter/enhancer element into close proximity with a specific OR gene, thus allowing its expression[179]. To test this hypothesis, two groups isolated mature OSNs and transferred their nuclei into enucleated oocytes; these were then used to produce chimeric or clonal mice that carried the genome of the specific OSN used for the transfer. The created animals were normal, able to produce a fully developed olfactory system with mature OSNs; these expressed several ORs and projected to multiple glomeruli[180, 181]. Furthermore, nuclear transfer experiments were performed using the nuclei of OSNs specifically expressing the ORs M71[181] or P2[180]; the resulting animals also expressed multiple different ORs and innervated all glomeruli. Analysis of the *M71* or *P2* loci revealed no signs of recombination or any sequence alterations in comparison to wild-type animals[180, 181]. Therefore, it was concluded that irreversible DNA recombination does not account for the expression of a single OR gene in OSNs.

A third hypothesis suggested the existence of a locus control region (LCR) capable of interacting with the promoter of a specific OR gene to activate transcription[179]. The availability of a single LCR in the genome would ensure singular expression. This theory gained momentum when an enhancer element was identified that could work as an LCR. It had been previously observed that a large YAC, containing hundreds of kb upstream of the MOR28 cluster was able to produce monogenic and monoallelic expression in OSNs, when inserted as a transgene. However, truncated versions of the YAC showed no expression whatsoever. Analysis of the sequences upstream the *MOR28* (*Olfr1507*) gene revealed a 2.1 kb segment that is conserved between the mouse and human genomes and when missing from the YAC, expression was abolished. Given the homologous nature of this sequence it was termed the H region[182]. The H region lies 75 kb upstream of the MOR28 cluster, which contains seven genes: *MOR28*, *MOR10* (*Olfr1508*), *MOR83* (*Olfr1509*), *MOR29A* (*Olfr1510*), *MOR29B* (*Olfr1511*), *MOR30A* (*Olfr1512*) and *MOR30B* (*Olfr1513*). Between *MOR83* and *MOR29A* there is a T cell antigen receptor gene; the first three ORs are expressed in the ventral part of the MOE while the last four are found more dorsally[183]. Within the 2.1 kb, there are 124 bp that are necessary and sufficient for the element to be able to induce expression; this is termed the H core. It contains three homeodomain binding sites and one O/E-like sequence; mutation of these sequences abolishes the enhancer activity[184]. When the H element

was attached to the *MOR28* sequence and inserted as a transgene, robust expression was observed in the MOE, whereas *MOR28* alone was never expressed, consistent with the results from the truncated YACs. These results led to the suggestion that the H region was an enhancer but, furthermore, that it could be an LCR[182]. Since then, the H region is also referred to as the H element or H enhancer[183].

To test whether the H element was able to regulate expression of OR genes in other clusters and chromosomes, chromosome conformation capture (3C) experiments were performed. In this methodology, the chromatin is treated with paraformaldehyde to crosslink the proteins and DNA that are interacting in the cell nucleus; then the DNA present in these complexes can be recovered and sequenced, to identify the sequences that were in close proximity. 3C experiments directed at the H element revealed that several different OR genes, from many chromosomes, were interacting with the enhancer. The most common interaction was with *MOR28*, followed by *MOR10*, the two OR genes closest to the H element; but at least other 20 ORs were identified as interactors. These experiments were validated by DNA and RNA-FISH, showing colocalisation of the H element with the *M71* or *M50* (*Olfr*6) OR genes and their corresponding RNA. Given these results, it was postulated that the H element was an LCR able to interact in *trans* with a single OR gene and activate its transcription[185]. Such a model was an attractive explanation for the monogenic expression pattern of OR genes. However, it was rapidly disproved by two groups which deleted the H element[183] or the H core[184], and showed that only the ORs from the MOR28 cluster were affected, while the rest of the receptors tested, either from the same or different chromosomes, were expressed at similar levels than in wild-type animals[183, 184, 186]. Importantly, in heterozygous animals only the OR genes from the MOR28 cluster on the same chromosome as the remaining H element were expressed, suggesting that the enhancer is not able to interact in *trans* to rescue the cluster in the other chromosome. Therefore, the H region was reassessed as a *cis* regulatory element able to influence the expression of the MOR28 cluster only[183, 184].

A similar region to the H element was later on identified, between the *P3* (*Olfr713*) and *P4* (*Olfr714*) OR genes. It is a 306 bp segment that shares 70% identity with the *P3* promoter; it is therefore named the P element[91]. It is situated near the end of a cluster of 24 OR genes, followed by other 43 receptor genes 670 kb downstream. To test if this sequence had similar properties to the H element, it was deleted and the expression of 577 different OR genes was assessed. Only nine genes were differentially expressed compared to wild-type animals, and all resided in the same cluster as the P element. In heterozygous animals, the single copy of the P element could not rescue

the expression of the genes in the other chromosome, implying that its activity is in *cis* only. Importantly, the differential expression observed through expression estimates, was validated with *in situ* hybridisation cell counts, meaning that the differences in expression were due to a change in the number of cells expressing those OR genes. These results were extended to the H element as well. Therefore, both the H and P elements influence the probability with which an OSN chooses a particular OR gene from those in the cluster they regulate; they do not, however, influence the transcriptional activity of the promoters themselves[186].

With the advent of genome wide technologies and the growing body of evidence on the importance of epigenomic regulation on gene expression, it was possible to identify further putative enhancers controlling other OR clusters. Genome wide chromatin immunoprecipitation coupled with sequencing (CHiP-seq) against common histone modifications was performed in the MOE. Both the H and P elements showed particular positioning of different histone modification marks in and around the enhancer sequence. This pattern was then used to search for similar intergenic regions along the genome. After several filters, 35 putative regulatory elements were defined, with an average distance of 35 kb to the nearest OR gene. Several of these sequences were able to drive expression of a reporter gene, supporting their role as enhancers. A few were used to create transgenic mice and showed widespread expression in the MOE and MOB, similar to what was observed with the H element. Finally, evidence of their possible involvement in regulating OR expression came from the deletion of one such enhancer, which led to the downregulation of the OR genes in the nearby cluster[187]. All these recapitulate what has been observed for the H and P element, suggesting that these sequences could be indeed enhancers involved in regulating different OR clusters; however, the definite proof of their influence on OR expression has been confirmed for only one of the 35, so all the other remain as putative candidates.

Analogous experiments to the 3C strategy used to identify which sequences the H element interacts with were performed with this new set of enhancers. These revealed that 32 out of the 35 sequences are frequently found in close proximity with the other enhancer elements; some are promiscuous and interact with many while others are more specific. Such interactions were confirmed by DNA-FISH experiments. Interestingly, these putative enhancers have binding sites for BPTF, a histone binding component of a chromatin remodelling complex. Knockout of *Bptf* resulted in the loss of OR expression and fewer interactions between pairs of enhancers could be detected by DNA-FISH, suggesting that the interactions were abolished. Based on all these, the authors proposed

a model whereby each enhancer element is necessary only for the expression of the OR genes in its nearby cluster, but interactions in *trans* that bring together many of these enhancers allow the robust expression of the chosen OR. In this scenario, knock-out of a single enhancer element wouldn't have an effect on the expression of the majority of the receptor repertoire[187].

1.2.1 *Cis*-acting elements influence OR expression.

In an effort to understand how is OR expression regulated, several groups reasoned that the use of transgenes could shed light into which features are fundamental to recapitulate the characteristic elements of OR expression: it should be monogenic, monoallelic, in a punctate pattern restricted to a subregion of the MOE and axons with the same OR should coalesce into a particular set of glomeruli[188]. A 9.4 kb construct containing the *MOR23* (*Olfr16*) OR gene was used as a transgene; the gene contains two 5' non coding exons, followed by the CDS contained in a single exon. The construct contained 400 bp upstream of the putative transcription start site (TSS) and 1.7 kb downstream of the stop codon. When randomly inserted in the genome, expression could be detected specifically in the MOE, in a monogenic, monoallelic punctate pattern, that was restricted to the zone of the epithelium where the endogenous gene is expressed. Furthermore, the axons of the OSNs expressing the transgenes co-converged with the axons of OSNs expressing the endogenous *MOR23* gene into one medial and one lateral glomeruli[88]. These results were recapitulated for the *M71*[88] and *MOR262-12* (*Olfr157*)[189] genes. In some cases, however, certain transgenic lines expressed the transgene in an aberrant pattern in the MOE, extending to other zones for example. This had a concomitant effect on the projection to the MOB and resulted in generation of additional glomeruli in shifted positions[88, 189]. The variability in the expression pattern for the same transgene in independent mouse lines probably stems from differences in the insertion locus in the genome. Nonetheless, it was remarkable to observe such tightly regulated expression with these small constructs, which were called minigenes[88]. Expression was achieved with only 405, 161 and 358 bp of 5' sequence for the *MOR23*, *M71* and *MOR262-12* genes respectively[188]. For the *MOR262-12* gene, it was confirmed that the insertion sites were not on the same chromosome as the endogenous OR gene, ruling out the possibility that the remarkable recapitulation of expression was due to insertion around the same locus[189].

To further delineate which sequence elements are necessary to obtain such patterns of expression, sequential deletions were made on the *MOR23* construct. Deletion of

the second intron had no effect; further deletion of the first intron, however, resulted in expression in three zones and the generation of additional glomeruli. Then, 1.4 kb of 3' region were deleted, which showed little effects. However, deletion of the 395 bp upstream of the TSS resulted in no expression of the transgene at all. Sequence analysis of the upstream region of the construct revealed the presence of six motifs for the O/E family (*Olf-1*, *Ebf1*) and a homeodomain (HD) protein binding site; deleting the upstream region of the TSS removed four out of the six O/E sites, suggesting that these are important for the expression of the transgene[88]. Similar motifs were identified on the promoters of the other two genes[88, 189]. To test the function of such motifs, deletion experiments were performed on the *M71* minigene. Shortly before the TSS there is an HD and an O/E motif in close proximity to each other. A transgene that loses all the upstream sequences except the 161 bp containing these two motifs is expressed in the expected pattern; but deletion of part or all of this 161 bp region results in loss of expression. To demonstrate that expression depends on these motifs, they were mutated either on their own or in combination. Mutation of either site resulted in OSNs expressing the transgene in a region ventral to the endogenous expression zone and the loss of the HD binding site also lowered the number of positive OSNs. When both sites were mutated together, the expression was completely abolished. Interestingly, when the same mutations were introduced into the endogenous *M71* promoter, the expression was drastically reduced and ventralised, but not completely lost, suggesting that other factors also contribute to expression regulation[190]. On the other hand, when a segment of 19 bp, that is conserved between the H and P element and contains an HD binding site, was inserted nine times into the *MOR23* minigene, the frequency with which the transgene was expressed was greatly increased, while maintaining the correct monogenic and zonal expression, and without altering the glomerular projections[188]. Minigenes have been successfully constructed for other genes, like *P3* and both the mouse and human *M72* receptors. All share conserved sequences in a short region upstream of the TSS, that are necessary for expression of the transgenes[188].

The number and arrangement of the different transcription factor (TF) binding motifs are variable between the promoters of different OR genes; besides the HD and O/E sites, other motifs are recurrently found conserved. It has been observed that receptors that share similar expression patterns in the MOE, like the receptors expressed in the patch area, have characteristic blueprints of such motifs. Similarly, the class I ORs have been proposed to have a distinct organisation of their promoter sequences[191]. However, extending this type of analysis to the whole repertoire wasn't possible initially,

since most of the annotation of OR genes has been done by homology searches with a small number of experimentally validated genes[62], and most include only the exon containing the CDS. What's more, evidence suggested that many of the receptors had 5' non coding exons that frequently presented alternative splicing[62, 192], and several polyadenylation signals, resulting in distinct 3' isoforms[62]. Separate groups used different technologies to map the TSS for several hundred mouse OR genes, which allowed a more comprehensive analysis of the receptors' promoters[192–194]. Usually, the promoter was situated several kb upstream of the CDS [192, 193]. Consistent with previous studies, the consensus TF binding sites in most OR promoters were O/E-like and HD sites; the O/E-like sites tend to cluster in the 50-150 bp upstream the TSS while the HD sites are preferentially found within the 100-150 bp 5' to the TSS, sometimes extending up to 600 bp[192–194]. A rigorous motif search analysis scored the O/E motif and HD sites specific to *Lhx2* and *Emx2* the highest[195].

The role of both *Lhx2* and *Emx2* in OR expression regulation has been confirmed. A yeast one-hybrid assay against the HD site in the *M71* promoter captured both proteins. *Emx2* is expressed homogeneously in the MOE while *Lhx2* is found predominantly in the basal layer of progenitor cells and its expression decreases more apically[196]. The *Lhx2* knockout is lethal; E16.5 embryos lack an MOB[196] and have very few mature OSNs, which are restricted to the dorsomedial region of the epithelium[197, 198] and express class I ORs, though at reduced levels and in fewer cells compared to control animals[198]. The expression of class II ORs is completely lost in the mutants. The expression of markers of progenitor cells is normal but as these differentiate into OSNs, there is 3.5 fold increased apoptosis and a failure to transit into mature OSNs[196, 197]. Knocking out *Emx2* also is lethal but the E18.5 embryos possess an overall normal MOE except it is thinner than controls; this is the result of a loss of almost half of the OMP^+ population of mature OSNs, while immature neurones are unaffected. Interestingly, the expression of ORs is generally downregulated but a few receptors are expressed at higher levels. These are expressed in a much greater number of OSNs, indicating that the lack of *Emx2* alters the frequency with which these ORs are chosen. This might be the result of losing the ability to choose the ORs that do depend on *Emx2*, thereby freeing OSNs, that otherwise would be committed, to the rest of the repertoire[199].

1.2.2 Early-bird-gets-the-worm paradigm of OR expression.

A different use of transgenes was devised to unravel how are ORs expressed during the maturation of OSNs. Nguyen et al.[200] created a transgene containing the CDS

of an OR under the control of the synthetic TetO promoter, that is activated with a tetracycline transactivator (tTA). Compound heterozygotes that expressed tTA under the control of the *Omp* promoter, showed widespread expression of the activator in all mature OSNs; yet, the transgene OR was expressed in only 10 to 30% of the neurones. Therefore, OR genes are vulnerable to silencing even under the control of an artificial promoter. Importantly, the cells that expressed the transgene obeyed the rules of monogenic and monoallelic expression, indicating that the CDS alone is also able to silence the endogenous receptor repertoire. However, when the expression of tTA was driven by the promoter of *Gγ8*, which is expressed in immature OSNs prior to endogenous OR expression, the transgene was no longer silenced. In adult animals, however, since mature OSNs had turned off the expression of *Gγ8,* expression of tTA ceased, and the transgene could only be detected in regenerating neurones. To maintain expression in adults, tTA had to be controlled by both the *Gγ8* and the *Omp* promoter; in this situation, the great majority of OSNs expressed the transgene OR. These data underpin an "early-bird-gets-the-worm" hypothesis, where the first OR to be expressed manages to avoid silencing by the rest of the receptor repertoire. Importantly, the silencing is dependent on the CDS and not the promoter, but this applies only to ORs and not other GPCRs. Furthermore, the signalling cascade activated by ORs is not required for this process, since a mutant OR that is unable to couple to $G\alpha_{olf}$ is still expressed monogenically[200].

A similar approach was used to study the transcriptional permissiveness of the *P2* gene, on its native locus. The TetO promoter was inserted upstream of the TSS of the endogenous *P2* receptor. The expression of tTA was driven by the *Omp* promoter and this resulted in increased P2 expression, in a zone-dependent manner. The greater increase was observed in P2's native zone and the effect faded as distance increased. Based on this, the authors proposed that zonality is achieved by differential chromatin organisation, whereby loci of receptors from different zones are made inaccessible to the transcription machinery when cells are outside their expression domain. In this model, the TetO promoter would have no influence on the expression of *P2* outside its zone, because tTA wouldn't be able to access it. This graded silencing of the TetO promoter was independent of the CDS, since the same results were obtained in an analogous animal that lacked the receptor's ORF. The authors further hypothesised that the frequency of choice of a particular OR gene is dependent on the permissiveness of its locus. To test this, the expression of *P2* was allowed to be activated by tTA, and then doxycycline (dox) was added, which blocks its activity. Despite the lack of the transactivator, *P2*

expression persisted, now from the endogenous promoter.

Moreover, when dox was administered during embryogenesis and into the first five days after birth, the number of *P2* expressing cells was dramatically reduced; and if the tTA was only active from postnatal day 30 (P30) to P60, mice were indistinguishable from controls with no induction. Therefore, activation of the *P2* locus is possible only during a short window in the maturation process of OSNs; once they have chosen an OR, the tTA is not enough to activate *P2*, even in cells from its epithelial zone. Interestingly, the silencing of the artificial promoter extends to ensure receptors are expressed monoallelically. In a compound heterozygote where each allele was labelled with a different reporter, both under the TetO promoter, only 3% of the OSNs showed co-expression of the two alleles. These 3% of OSNs were situated in the most basal neuronal layers suggesting that they were still young, newly-differentiated neurones. Therefore, despite having the ability to express both alleles, cells chose only one. The authors hypothesised an asymmetry between the alleles, making one more likely to be activated and then able to suppress the rest of the repertoire, including the other allele[201].

Nguyen et al.[200] observed the same with the TetO-*P2* transgene. In this case, the expression of tTA was controlled by the endogenous *P2* promoter; therefore, expression of the transgene meant that both the endogenous *P2* and the transgene *P2* were produced in the same cell. However, this was observed in less than 2% of the OSN population and always in basally-located neurones[200]. In this regard, Chess et al.[72] described asynchronous replication of the two alleles of an OR gene when they identified the monoallelic character of OR expression. Asynchronous replication of OR genes has been observed as early as embryos that have passed the blastula stage and this is maintained through cell divisions. Through differential epigenetic marking, the allele that is replicated first becomes more available for expression[202]. Therefore, during OSN generation, one allele is already set in a more permissive state than the other.

Epigenetic regulation plays a fundamental role in allowing OR expression to occur in such a peculiar way. The OR loci in the MOE are characterised by chromatin modifications H3K9me3 and H4K20me3, which are characteristic of constitutive heterochromatin (pericentromeric and telomeric repeats). This type of heterochromatin is highly condensed throughout the cell cycle and is maintained this way during development. Analysis of the positioning of both of these marks revealed a clear concentration around OR and VR loci, forming a macrodomain that extended throughout the receptor cluster. Analysis of horizontal basal cells revealed that OR loci were marked by H3K9me2, which is commonly found in facultative heterochromatin and is plastic throughout de-

velopment. However, in precursor and immature OSNs, the constitutive heterochromatin marks were already laid, indicating that this occurs before OR expression[203]. Moreover, the compacted chromatin containing the OR genes clustered into an average of five highly compact foci per OSN[204]. This 3D organisation in the nucleus might facilitate the observed interactions in *trans* of all the identified enhancer elements[187]. Notably, the allele that was expressed in a given OSN didn't colocalise with these foci and was instead located nearby euchromatin and active PolII regions[204].

The compaction of OR loci into foci is dependent on downregulation of LBR, a nuclear envelope protein that interacts with heterochromatin. LBR is highly expressed in progenitor cells but decreases in abundance with differentiation. When its expression was forced in mature OSNs, the foci were lost and the OR genes became sensitive to DNase I cleavage, indicating a decompaction of the chromatin. Additionally, mature OSNs expressed several ORs but at low levels[204]. This suggests that the gained accessibility to the OR repertoire allowed the expression of several receptor genes and, also, that the loss of enhancer interactions resulted in the loss of robust OR expression[187, 204]. Based on all these data, a model emerged whereby the basal state of the OR repertoire in maturing neurones is of widespread repression and singularity is achieved by de-silencing a single receptor[203].

In order to achieve expression of a particular OR allele, it is necessary to erase its silencing modifications and mark it for transcriptional activation instead, with H3K4me3 [203]. LSD1 is the only protein capable of catalysing the demethylation of both H3K9me2 and H3K4me2. It was observed that if this protein was knocked-out before the receptors were activated, there was a widespread loss of OR expression and mature OSN markers; but if it was knocked-out during or after OR gene activation there were no observable effects. Therefore, its activity is necessary to initiate OR expression, but not to maintain it. Consistently, the activated OR alleles, showed activity of LSD1 on their promoters, directly linking this protein to the desilencing mechanism[205]. This process is tightly regulated, as the activated allele was found robustly marked with the activating H3K4me3 modification, but neighbouring ORs retained their heterochromatin marks[203]. In the early *Lsd1* KO animal, which lacks OR expression and mature OSN markers, the introduction of a TetO-*M71* transgene was able to restore *Adcy3* expression. The presence of *Lsd1* was shown to be mutually exclusive with *Adcy3*, indicating that once a neurone expresses an OR and *Adcy3* expression is induced, *Lsd1* is shut down. This was demonstrated in an *Adcy3* KO, where OSNs that first chose a particular OR, then went on to choose a different one and kept on switching indefinitely.

Therefore, *Lsd1* is required to desilence an OR allele by demethylating its H3K9me2 (repressive); however, if still expressed once this allele is activated, it is also capable of shutting it down by demethylating its H3K4me2 (activating) and the process can be repeated indefinitely without ever achieving stable OR expression [205].

1.2.3 Negative feedback ensures singularity.

Several studies using transgenes used the promoter elements of OR genes to drive the expression of reporters, without including an OR CDS. Repeatedly, it was observed that the OSNs that initially chose this deletion allele for expression, went on to choose other receptors and the resulting neurones innervated multiple glomeruli[85, 91, 182, 206, 207]. This indicates that the CDS is necessary to stop the activation of other OR alleles. Importantly, a transgene containing the CDS of the *M4* (*Olfr63*) OR, but lacking the start codon, gave the same phenotype as the transgene lacking the whole CDS. Based on this, it was proposed that singular expression of OR genes was achieved by an OR protein-mediated feedback mechanism, rather than by restricting only one active promoter per OSN[206]. The ability of an OSN to activate two different OR promoters if the first yields no functional receptor is highly advantageous, given the high proportion of OR pseudogenes. Without this mechanism, a considerable number of OSNs would be stuck without a functional OR able to sense odorants, which would be a costly waste of resources.

Given that several promoters can be activated in the same cell, the question arises of how stable is OR expression. To answer this, Shykind et al.[207] performed a series of elegant lineage-tracing experiments. The *MOR28* endogenous receptor was engineered to express, along the receptor, the Cre recombinase; this mouse was crossed into a background containing a floxed reporter gene that, once activated, would be ubiquitously expressed. Therefore, all those OSNs that choose the *MOR28* promoter for expression at any given time, would be permanently labelled. Interestingly, when the labelled cells were checked by *in situ* hybridisation for *MOR28* expression, only 90% of the cells were positive. Further analysis of the 10% of cells negative for *MOR28* revealed that they expressed other ORs typically expressed in the same zone as MOR28, including the other *MOR28* allele. These other ORs were observed with frequencies similar to those with which they are normally chosen in the epithelium. Thus, these tracing experiments indicate that a fraction of OSNs extinguish the expression of the first OR they choose and select another receptor from those available for expression within the epithelial zone where they are located. The choice mechanism of this second OR is in no way

biased towards particular genes; instead it reflects the frequencies of choice normally observed[207]. Similar experiments were performed with the *MOR28* promoter driving Cre expression, but lacking the CDS. In these cells, the same phenomenon was observed, except now all labelled cells switched to express another OR. Interestingly, all these cells shut off the *MOR28* promoter since Cre could no longer be detected. What's more, this was observed for naturally occurring pseudogenes, that could be detected in young animals but were greatly reduced in number in adults, suggesting that the expression of non-functional OR proteins is extinguished with time[207].

OR gene switching is fundamental to avoid committing an OSN to express a non-functional OR. This process has also been documented to play a role in the zonal restriction of at least one type of receptors: the OR37 genes. These receptors are expressed in the patch, instead of the more canonical zonal expression pattern. Similar tracing studies as those described above were performed with a particular member of the OR37 subfamily, *OR37C* (*Olfr157*). Labelled cells were identified in an area larger than just the patch region; however, the *OR37C* RNA could only be detected in the cells located in the appropriate location, whereas all the others expressed different ORs. This indicated that a larger population of cells initially activated the *OR37C* promoter for expression but, when located outside the patch region, switched to a different gene[208].

So what could be mediating this protein-dependent feedback mechanism to ensure stable and singular expression? Dalton et al.[209] hypothesised that, since ORs are one of the most abundantly expressed proteins in OSNs, initiation of their transcription could result in endoplasmic reticulum (ER) stress. If so, this could be an indication that OR expression had been activated. Cells are equipped with a sensing system that monitors the amount of unfolded proteins present in the ER. If the protein levels are too high, a series of events are set in motion –the unfolded protein response (UPR)– to decrease the ER load; these include the induction of chaperones to aid in protein folding and decreasing translation initiation events. A sensor of the UPR is the production of the nuclear isoform of ATF5 (nATF5). *Atf5* contains an upstream ORF that inhibits the production of nATF5, but when translation initiation is slowed down, ribosomes are able to assemble on the downstream ORF and produce the nuclear isoform. The authors showed that induction of expression of an OR gene was sufficient to induce expression of nATF5, suggesting that the UPR was involved in signalling the presence of an active receptor gene.

PERK is one of the proteins activated upon detection of unfolded proteins; indeed, OR expression induced PERK activity, which in turn phosphorylates eif2α, a translation

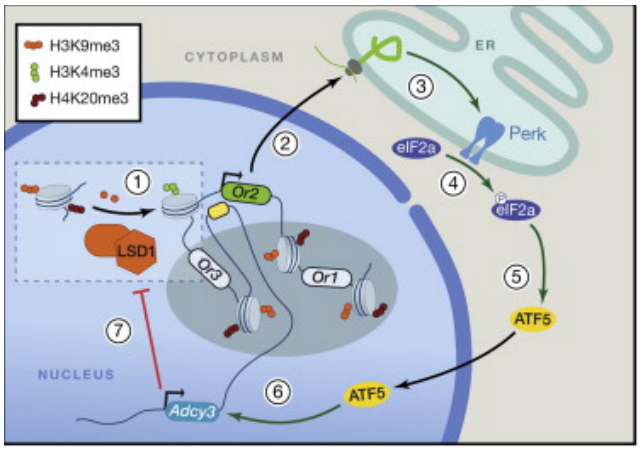


Figure 1.12 – A feedback mechanism ensures singular OR expression. Initially, the OR repertoire is silenced by condensation in foci (dark area in the nucleus) and marking with heterochromatin histone modifications (H3K9me3 and H4K20me3). (1) To activate OR expression, a single OR gene is desilenced by activity of LSD1. (2) The OR mRNA is translated and transported to the endoplasmic reticulum (ER), (3) where it activates the unfolded protein response (UPR) and activates the Perk signalling pathway. (4) This leads to the phosphorylation of eIF2a (5) and then the production of the nuclear isoform of ATF5. (6) In turn, ATF5 activates the expression of *Adcy3*. (7) Finally, ACIII shuts down LSD1 expression, ensuring no other OR genes are desilenced. Reprinted from [210], copyright (2013), with permission from Elsevier.

initiator factor, slowing down translation initiation and allowing the accumulation of nATF5. When ATF5 was knocked-out, *Adcy3* expression was dramatically lost, along with other mature OSN markers. In contrast, in *Adcy3* mutant animals, the expression of ATF5 was greatly expanded. Therefore, *Adcy3* is important in shutting down the UPR which is necessary to restore translation and allow the terminal differentiation of the neurones. At the same time, expression of *Adcy3* accompanies the downregulation of *Lsd1,* ensuring that other receptor alleles are not activated (Figure 1.12). A final elegant demonstration of the involvement of the UPR on eliciting the feedback signal of OR expression, involved the treatment of LSD1 KO animals –which are unable to activate OR expression– with tunicamycin, a drug that activates the UPR. This resulted in the expression of *Adcy3* and other mature OSN markers, suggesting that induction of the UPR can substitute for OR expression in eliciting the feedback mechanism[209].

Taking all these data together, Tan et al.[211] built a mathematical model that showed that singular OR expression in OSNs is determined by two parameters: the rate of OR activation and the latency to the negative feedback elicited by the activated allele. The most parsimonious model indicates that singularity is achieved by inefficient desilencing of OR alleles, probably at the stage of demethylating H3K9me3→H3K9me2, on which LSD1 can act. In this scenario, once a first OR locus is desilenced, the negative feedback mechanism is able to downregulate *Lsd1* before another demethylation event can occur. It is likely that the yet unidentified enzyme catalysing H3K9me3 demethyla-

tion is expressed at very low levels, greatly restricting its activity; this, however, is still yet to be proven[211].

1.3 Detection of odorants by olfactory receptors.

It took seven years from the discovery of the receptor proteins until the first specific OR-ligand interaction was identified. Zhao et al.[212] used an adenovirus carrying the rat *I7* OR gene coupled to GFP to infect the rat MOE. Around 1 to 2% of all the OSNs expressed the construct; the infection rate was not uniform across the epithelium and some areas had as many as 20% of all neurones infected. These regions allowed to perform EOG recordings upon stimulation with odorants. A panel of 74 ligands were chosen, with diverse molecular structures and odour qualities. Compared to controls, a significantly increased response was detected upon exposure to octyl aldehyde, otherwise known as octanal. This response was dependent on the expression of I7, as a vector containing GFP alone did not elicit an increased response. Further validation was obtained by whole-cell patch clamp recordings from single GFP^+ neurones. I7-expressing cells were responsive to other saturated aliphatic aldehydes with carbon chain lengths from 7 to 10 carbons (C_7-C_{10}). No response could be elicited with C_6 hexanal whatsoever, but clear activation was achieved with C_7 heptanal, showing the remarkable ability of the receptor to discriminate between these two. Additionally, other aliphatic compounds with varying functional groups failed to produce a response. These experiments thus showed that the identified multi-gene family of ORs indeed were able to bind to odorants and generate an electrical response[212].

Previous studies utilised calcium imaging of dissociated OSNs to study their responsiveness to different types of ligands. These were informative on the properties of the neurones, but never proved direct receptor-ligand relationships, or the dependence of the response on the OR itself. Nonetheless, it quickly became apparent that particular OSNs are tuned to discriminate different molecular characteristics of odorants. For example, a study exposed dissociated OSNs to fatty acids or aliphatic alcohols with varying hydrocarbon chain length. Some neurones were responsive to only one of the classes, indicating that the functional groups can be differentiated by some receptors. However, other OSNs were activated by both classes of odorants, but instead were highly selective for the length of the carbon chain. In all cases, the responses were greatest for one or two molecules, and the sensitivity decreased as the molecules became more dissimilar, requiring higher concentrations to achieve a response[213].

1.3.1 Combinatorial olfactory coding.

A different strategy to reveal specific OR-ligand interactions was used by the group of Linda Buck. In this case, dissociated OSNs were analysed by calcium imaging upon stimulation with a panel of aliphatic odorants with chain length from 4 to 9 carbons, comprising alcohols, carboxylic, bromocarboxylic and dicarboxylic acids. The responsive OSNs were then subjected to single-cell RT-PCR with degenerate primers for OR genes. 46 neurones activated with at least one compound were analysed and half of them yielded a PCR product. A subset of these were sequenced and each provided a different OR gene, with identity values ranging from 19 to 100% in the TM3-TM6 region, which contains the putative odorant binding site. Most of the activated neurones responded to several odorants, showing specificity for the carbon chain length. None responded to all four classes of molecules. Further, a single odorant activated several OSNs and, therefore, different ORs. Thus, olfactory responses depend on a combinatorial code, whereby a single odorant activates several receptors, each of which responds to several odorants (Figure 1.13). This strategy allows for the coding of an immense number of ligands. Interestingly, the set of receptors that respond to a particular odorant can be very similar or very dissimilar to each other, with identity values dropping as low as ~22%[69].

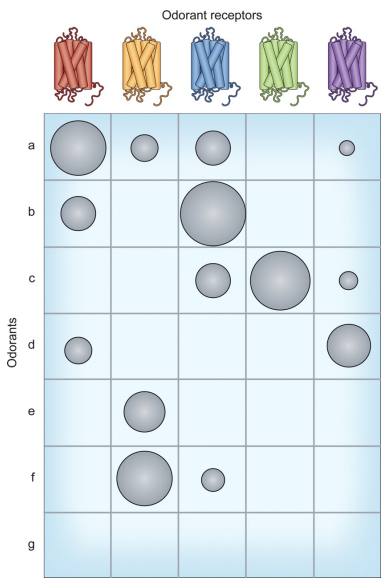


Figure 1.13 – **Combinatorial odour coding.** At the top are represented five different ORs. Below is a table of the responses of each OR to seven different odorants (a to g). The diameter of the circle indicates the magnitude of the response. Some odorants activate many receptors while others activate only a few. Also, some ORs respond to only one odorant while others can bind to several different compounds, with varying affinity. Each odorant elicits a particular pattern of OR activation. Reproduced from [83].

The same OR is able to respond to different ligands with varying sensitivities and the same odorant activates different receptors at distinct concentrations. When odorants are presented at higher concentrations, more ORs are active. This means that the receptor code for a specific odorant is dependent on its concentration and not only on the molecular structure[69, 214, 215]. It has also been observed that OSNs that all express the same OR can respond to the same odorant with sensitivities that span over two orders of magnitude; but if they respond to two different odorants with different affinities, the relationship between the two is consistent across the neurones. For example, OSNs that express the M71 receptor have been shown to respond preferentially to acetophenone, but also to benzaldehyde at higher concentrations. Different cells show very dissimilar affinities for acetophenone, but they are always more sensitive to acetophenone than to benzaldehyde[216].

The combinatorial code can be observed in the MOB also, at the level of individual glomeruli. The dorsal olfactory bulb is accessible with minimal surgical manipulation in live individuals and it can be imaged while animals are presented with different odorants. As observed at the receptor level, each odorant activated several glomeruli, and the same glomerulus responded to different stimuli. Each ligand elicited a particular pattern of activation that was different even with small changes in the chemical structure of the compounds. Also, increasing concentrations of the stimulus resulted in the recruitment of additional glomeruli that were not activated at lower concentrations, exemplifying the differences in sensitivity for different ORs to the same agonist[217].

As OSNs expressing the same OR are scattered along a restricted portion of the MOE, the neurones that respond to a particular odorant are also dispersed and intermingled with non-responding cells. However, a much larger proportion of cells respond to a given ligand compared to the number expressing a particular OR. Also, responsive OSNs are not restricted to the epithelial zones of receptor expression; instead, they are found in both dorsal and ventral portions of the MOE[218]. Consistently, the receptor responses to its agonists are not dependent on the zone it is expressed. In the rI7→M71 mouse, the CDS of the *M71* mouse receptor was substituted for that of the rat *I7* gene, tagged with GFP. *I7* is normally expressed in the ventral region of the epithelium but in this transgenic mouse it is found in OSNs of the dorsal domain. Calcium imaging of the fluorescent cells revealed that they were responsive to octanal, but not to acetophenone, the ligand of M71. Therefore, the promoter controls the frequency and pattern of expression of the receptor in the MOE, but has no influence on its binding profile. Further, in a similar mouse that expressed the mouse *I7* receptor from the *M71* locus, the responses

recorded from these neurones were indistinguishable from those of OSNs expressing the endogenous *I7* gene, despite the drastic change in zonal expression[216]. The same was observed with the M72 receptor, either expressed from its endogenous locus or from the *S50* (*Olfr545*) receptor locus, which is not only expressed in a different epithelial zone but is a class I OR while *M72* is a class II[87].

The shift from a ventral to dorsal expression zone in the MOE results in a concomitant shift in the position of the corresponding glomeruli to the dorsal MOB. Direct imaging of the glomeruli in animals stimulated with octanal showed specific activation of the rI7→M71 glomeruli. Moreover, the glomeruli were innervated by mitral and tufted cells and recordings from these confirmed the stimulatory activity of octanal. What's more, tracing experiments on tufted cells innervating the dorsal glomerulus revealed connections specifically to the rI7→M71 medial glomerulus, thus linking the two halves of the bulb. Therefore, rI7→M71 expression results in the creation of additional glomeruli in an ectopic region of the bulb that, nonetheless, are responsive to the cognate ligand and form the appropriate functional circuitry to convey the olfactory information[219].

The rat *I7* receptor has been studied by many groups, using very different experimental systems. All the data is consistent with the initial observations: it responds to aliphatic aldehydes of backbone chains of 7 to 10 carbon atoms and octanal is its most potent agonist. A thorough characterisation of the molecular range of this receptor has been performed. Araneda et al.[220] screened 90 odorants, all closely related to octanal in their molecular structure, each with changes in the length of the hydrocarbon chain length, the functional group, side chain substitutions or the degree of saturation. This strategy demonstrated that the aldehyde group is necessary but not sufficient for activation. As noted previously, the size was tightly discriminated, and only compounds with 7 to 10 carbons were agonists. There was, however, tolerance for substitutions along the backbone, both in terms of double bonds and methyl groups and other substituents, especially after C_4. In the first three carbons, a double bond or a methyl group were well accommodated by the receptor, but the presence of both abolished binding[220].

A similar approach was applied to mOR-EG (*Olfr73*), an OR that responds to eugenol (EG). In this case, the benzene ring of eugenol was kept constant, while the functional groups around it were systematically changed. Screening with these compounds led to the identification of responses to vanillin and ethyl vanillin. Vanillin has an aldehyde instead of the allyl group found in eugenol[215]. Various other molecules with substituents at this position were also agonists at high concentrations, but charged groups were not tolerated[215, 221]. The complete removal of any of the functional

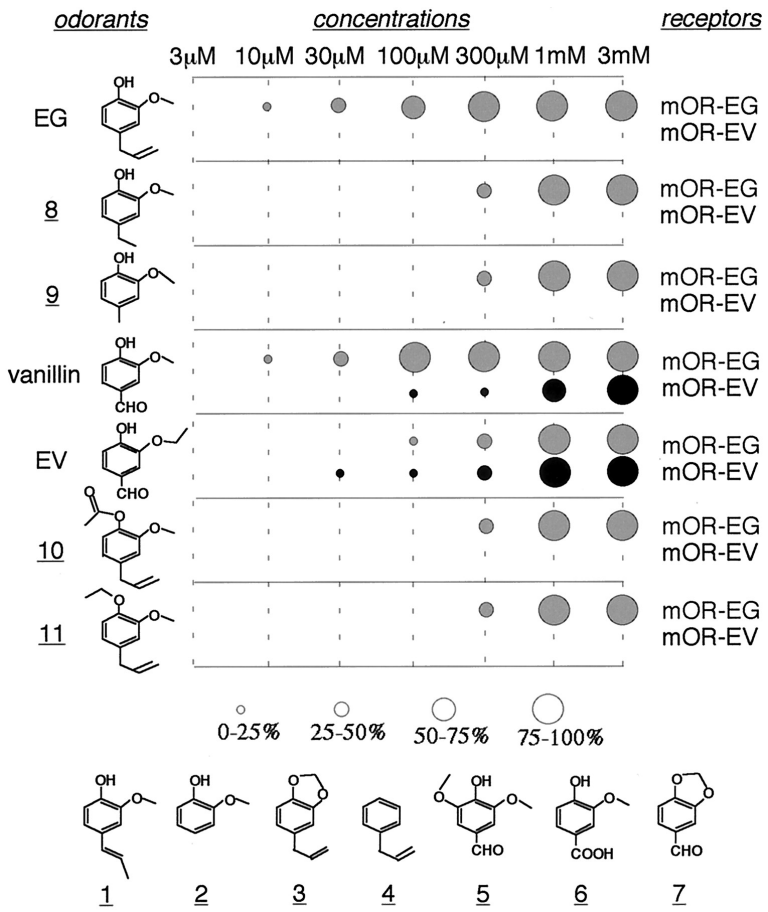


Figure 1.14 – Molecular range of ligands for the mOR-EG and mOR-EV receptors. Two mouse OR genes (mOR-EG and mOR-EV) were tested with a panel of odorants (at a range of concentrations) that are variations on the structure of eugenol and vanillin, the ligands of these receptors. Components 1 to 7 did not elicit any responses, while mOR-EG responded to 8-11 at high concentrations. Both ORs recognise vanillin but with differing sensitivities. Republished with permission of the Society for Neuroscience, from [215]; permission conveyed through Copyright Clearance Center, Inc.

groups along the benzene ring resulted in loss of activation. What's more, the creation of a stereoisomer by altering the double bond on the position of the allyl group also abolished activity, despite this being a subtle change in molecular structure[215]. At other positions, characteristics such as the size of the functional group were important, and an absolute requirement for activity was the presence of an oxygen attached to the benzene ring (Figure 1.14). In all, 22 compounds were able to activate mOR-EG, with affinity values spanning several orders of magnitude[221].

A reciprocal experiment was also performed, whereby OSNs responsive to octanal were screened to characterise the diversity of their activation profiles. While some neurones showed similar profiles to I7, others were tuned to discriminate between molecules that differed by just one extra double bond or were activated only by 8-carbon aldehydes.

Among all these octanal-responsive cells, there was a continuum of responses; some cells were broadly tuned to many different molecules and others were more selective to the point of responding to octanal alone[222]. However, it is important to keep in mind that how narrowly or broadly tuned a response profile appears depends on the stimuli presented; many of the neurones that seemed to be very specific to octanal are most likely activated by other molecules that were not tested.

With this in mind, a study selected a broad panel of odorants that comprised many different chemical properties, molecular structures and perceived odour qualities. These were grouped into 13 mixes containing structurally related compounds and used to screen dissociated OSNs by calcium imaging. The results agreed with the notion that most OSNs are narrowly tuned and responded to only one mixture. However, a small number of OSNs were also remarkably broad in their response profiles, showing activation by 5 to 12 different mixtures. Interestingly, it was also observed that different classes of odorants activate higher numbers of OSNs than others; for example, the mixture containing aldehydes elicited a response in 59% of all the screened OSNs, while the amines activated only 15%. Furthermore, disparities were evident with the different aldehydes themselves; octanal and decanal were able to activate around 40% of the aldehyde-sensitive neurones, while other aldehydes activated only 2.5% of the population. This could be the result of particular classes of ORs being more abundant within the OSN sample; however, within the screened cells many different response profiles were identified, suggesting the presence of distinct ORs[223], but this wasn't verified.

The restriction to bind particular molecular motifs is the result of key interactions between the odorants and specific amino acid residues within the binding pocket of the receptor proteins. A computational model of the structure of mOR-EG was constructed based on homology to another GPCR for which a 3D structure was available. Using the knowledge of the responses to the different agonists screened[215, 221], 10 amino acids in the binding region of the receptor were predicted to be functionally important for the interactions between the receptor and the various ligands. Mutation of some of these indeed changed the affinity for some molecules or completely abolished activity, demonstrating their role in ligand recognition[221]. The importance of particular amino acids has also been observed with naturally occurring variation. The mouse and rat I7 receptors differ in 15 amino acids, three of which are located in the putative binding site. In contrast to the rat I7 which is most sensitive to octanal, the mouse receptor is better tuned to detect heptanal and responds with less affinity to octanal. One amino acid change is sufficient to dictate the specificity for either odorant[224].

A similar situation arises with close paralogous OR genes within one species. For example, the *M71* and *M72* genes are 96% identical and both respond with high affinity to acetophenone[85, 216]. However, screening with a large number of odorants revealed that M72 is activated by at least 14 other compounds, some of which also activate M71; and M71 has robust responses to five other agonists, all of which elicit larger responses in M71 than M72. Therefore, despite the great similarity between the receptors, the odour profiles are clearly distinguishable between the two. Further experiments were performed with the *M71* receptor, with targeted changes at specific amino acids to reflect the ones found at the same positions in *M72*. In one such case, a single change resulted in an odour profile much more similar, but not identical, to M72 than M71. OSNs expressing the mutated receptor formed new ectopic glomeruli, different from both the endogenous receptors. A different mutation profile that altered 4 of the M71 amino acids to reflect the M72 sequence, led to intermingling of the axons with the ones from OSNs expressing the M71 endogenous receptor. The response profiles of the two types of receptors were similar but some differences were evident; this implies that two distinct response profiles can be mapped to the same glomerulus[87].

The importance of the physicochemical properties of the odorant molecules in the interactions with ORs is indisputable. However, molecular shape is also an important factor, and can be sufficient to alter the response profiles of the receptors. Enantiomers are mirror images of one another, that are non-superimposable. A remarkable example is that of (S)-carvone and (R)-carvone, which smell like caraway and spearmint respectively, despite their identical chemical properties. Imaging of the MOB while pairs of different enantiomers were presented to rats showed that some glomeruli were activated by both isomers while others were specific to one member of the pair. The activation patterns for enantiomers were more similar to each other than they were to other odorants, but they were still clearly differentiable[225]. Consistently with this, at the level of the OSNs, different subsets could be identified that were activated specifically by each isomer, and some that were responsive to both[218, 226]. Such OSNs contained a variety of different ORs[226].

1.3.2 Deorphanisation of olfactory receptors.

The identification of specific ligand-OR pairs has been slow and difficult. Until a few years ago, only a dozen rodent ORs had been deorphanised[2]. Several approaches have been developed, both *in vitro* and *in vivo*, with varying success rates. OSNs are the ideal system to express the OR of interest, given that they are equipped with all the neces-

sary machinery for the proper expression and trafficking of the receptors to the plasma membrane, as well as the components of the signalling pathway to detect activation. Infection with an adenovirus carrying a vector for overexpression of an OR was successful in identifying the first receptor-ligand pair and establishing its response profile[212, 220], and was also used to validate results obtained by other methods. An alternative strategy consisted on using dissociated –unmodified– OSNs in calcium imaging assays. This technique consists in loading the neurones with fura-2, a calcium indicator, that reports the intracellular concentration of Ca^{2+}. Since binding of ligands to ORs results in influx of calcium into the OSN, it is an efficient way of measuring OR activity. The activated OSNs can then be subjected to single-cell RT-PCR with degenerate primers for ORs, to identify the specific receptor that has shown a response[69, 226]. With this strategy Touhara et al.[227] identified the *MOR23* OR, after screening cells with lyral, and then corroborated the specificity of the interaction by the adenovirus technique. However, PCR reactions on single cells fail very often and the possibility of contamination is very high, which makes it necessary to demonstrate the interaction in an alternative system.

The generation of several mouse lines with ORs tagged with reporter genes facilitated the identification of the population of cells expressing a given OR. These cells could then be screened by calcium imaging with the *a priori* knowledge of the receptor being interrogated. This strategy was used to uncover the binding of acetophenone by M71 and M72[85, 87, 216]. An alternative strategy exploited the marking of the glomeruli innervated by the tagged receptors, to perform *in vivo* imaging of the MOB directly[228]. This methodology was used to deorphanise the MOR29A and MOR29B receptors, which are 95% identical at the protein level. Specific signals could be identified upon stimulation with aromatic odorants with phenyl ether groups; further studies on dissociated OSNs by calcium imaging showed that both receptors were activated by guaiacol and vanillin. However, some of the compounds that elicited responses in the MOB failed to activate the OSNs *in vitro*[229]. Such discrepancies between systems had been observed before[24, 228] and probably stem from the fundamental differences in which stimulation is performed: in the *in vivo* methods, odorants are delivered in the vapour-phase and with the mucus layer intact while in the other systems stimuli are delivered in an aqueous solution.

More recently, the importance of the nasal mucus was directly tested. Nagashima and Touhara treated the mouse mucus with different odorants and studied its composition after five minutes, by gas chromatography and mass spectrometry. Strikingly, though perhaps not that surprising, some molecules were found to be rapidly metabol-

ised. For example, 80% of the added benzaldehyde was converted to benzyl alcohol and benzoic acid and more than 90% of the acetyl isoeugenol was transformed into isoeugenol. Overall, aldehydes and acetates were readily decomposed into other compounds by the enzymatic activity of the mucus while the alcohols, thiols and ketones tested so far have not shown any signs of conversion. Importantly, behavioural tests were used to demonstrate that animals with intact mucus were not able to discriminate between the initial odorant and its subcomponents; whereas animals that were treated with inhibitors for the enzymes mediating the reactions were capable of distinguishing the compounds. Thus, the enzymatic transformation of the initial odorants occurs fast enough to affect the ligands that reach the OSNs and, consequently, the responses elicited[24].

The restriction to the dorsal MOB for *in vivo* imaging could be circumvented by replacing the coding sequence of dorsally expressed ORs by those expressed more ventrally, thus shifting their glomeruli to the accessible dorsal bulb. However, all the strategies involving the creation of transgenic animals are slow, expensive and low throughput[230]. Other approaches combining several of the above mentioned techniques have been used to identify the ORs responding to a particular odorant. For example, Oka et al.[228] exposed mice to eugenol, methyl isoeugenol and isovaleric acid while recording calcium responses on the glomeruli of the dorsal MOB. Then, used retrograde labelling to identify the OSNs innervating the activated glomeruli and, from them, performed RT-PCR to isolate the responsive ORs. Shirasu et al.[231], however, were unlucky and unable to identify muscone responsive glomeruli on the dorsal MOB, so they performed unilateral bulbectomy to allow the visualisation of the medial bulb. This is still a restricted view of a subset of the glomeruli but, in this case, revealed one to three glomeruli that responded upon stimulation with muscone; these were highly specific to musk-like compounds and failed to be activated by a varied range of other molecules. The specific ORs were then identified by a combination of retrograde labelling and calcium imaging of dissociated OSNs, followed by RT-PCR[231].

A final *in vivo* strategy was developed by McClintock and colleagues[232]. They utilised a calcium and zinc binding protein, S100a5, to drive the expression of GFP. This gene is transcribed in an activity-dependent manner[97] and, therefore, OSNs that are activated by an odorant become labelled. Targeted mice were exposed to eugenol or muscone, and their GFP population of OSNs was compared to that of animals treated with vehicle alone. Microarray assays were used to identify the ORs that were enriched after exposure to the odorants; this yielded three eugenol- and five muscone-responsive ORs. The receptors for eugenol had all been previously identified though mOR-EG

was not in the list. The muscone receptors contained the OR identified by Shiratsu et al.[231] plus other candidates, one of which was confirmed in an *in vitro* heterologous expression system[232]. The lack of signal for the best characterised eugenol receptor could stem from the high background noise level of microarrays; coupling this system to more sensitive transcriptional profiling methodologies, such as RNAseq or NanoString nCounter, could improve the results. Such a methodology has the potential to uncover a more complete catalog of all the receptors activated by a particular ligand in a realistic *in vivo* delivery setup. However, it represents a low-throughput, timely and expensive strategy for deorphanisation.

The most popular strategy of all is, undoubtedly, the use of heterologous systems to express the ORs and then directly screen for activity-induced responses. This strategy is fast, cheap and suitable for parallelisation to screen many receptors against many ligands at the same time. However, it has proven difficult to successfully express ORs in cells other than OSNs. Most approaches have used human embryonic kidney (HEK293) cells transfected with a construct carrying the OR and a Gα_{15} generic G-protein subunit for coupling. Additionally, a reporter such as luciferase is included, under a cAMP response element (HEK293 cells endogenously express an adenylyl cyclase), to visualise the activation of the OR. Alternatively, a calcium imaging approach with fura-2 has also been implemented[230]. The trouble with expressing ORs in heterologous systems is that most receptors never reach the plasma membrane. Rhodopsin, a GPCR that is part of the class of receptors most closely related to ORs, could successfully be expressed and localised to the membrane of HEK293 cells. Therefore, Krautwurst et al.[224] fused the 20 N-terminal amino acids of this protein (Rho tag) to the OR constructs, which enhanced proper expression of the receptor protein at the plasma membrane. Using this strategy, they were able to identify receptors that responded to carvone, citronellal and limonene. A different approach consisted of adding a cleavable influenza hemaglutinin signal (IHS) sequence at the beginning of the OR CDS; this was similarly successful for the expression of the receptor responsive to 2-heptanone and other 2-ketones[233]. However, addition of these tags was not enough for proper expression of other receptors.

Tracing studies revealed that, in HEK293 cells, the translated OR proteins were retained in the ER and failed to translocate to the Golgi apparatus, probably because they were misfolded. In the ER, the receptors were polyubiquitinated and targeted for degradation by proteasomes, or formed aggregates that were degraded by autophagy [234]. Therefore, several groups searched for chaperones or other accessory proteins present in OSNs that were lacking in HEK293 cells. This led to the dicovery of receptor-

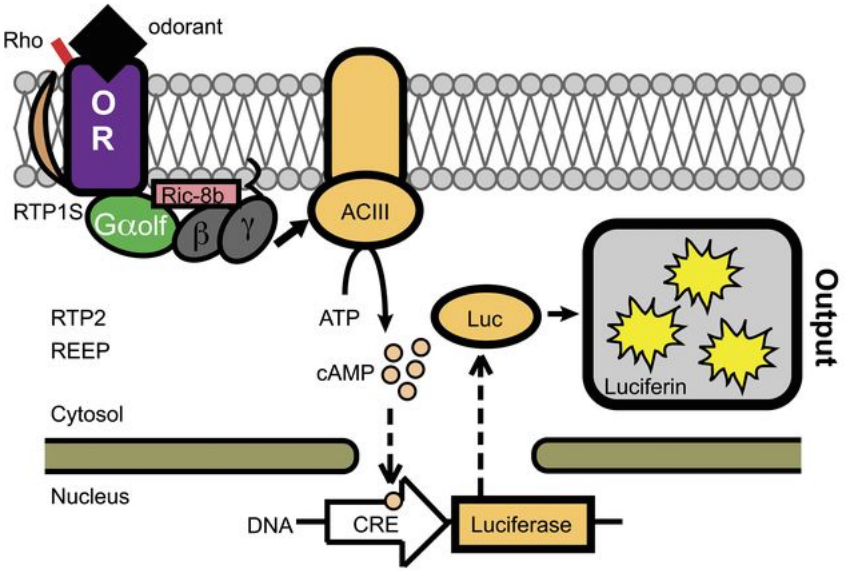


Figure 1.15 – *In vitro* expression of ORs in Hana3A cells. HEK293 cells stably expressing $G\alpha_{\text{olf}}$, RTP1S, RTP2 and REEP1, called Hana3A cells, are able to express a construct containing an OR gene and translocate it to the membrane with better efficiency than other systems. A luciferase protein is under the control of a cAMP response element (CRE). Upon activation of the OR with a ligand, ACIII is activated and produces cAMP, which in turn activates the transcription of the luciferase and the production of luminescence. Figure from [230].

transporting protein (RTP)1 and RTP2, as well as the receptor expression-enhancing protein (REEP)1. All these are specifically expressed in OSNs and are able to interact with OR proteins. When co-transfected into HEK293 cells, the ORs were translocated to the plasma membrane and luciferase signals could be recorded[235]. To facilitate deorphanisation tasks, a HEK293 cell line was established that stably expresses $G\alpha_{olf}$, RTP1, RTP2 and REEP1, and was called Hana3A[235]. Further studies later identified a shorter isoform of RTP1, named RTP1S, that is much more potent in promoting OR expression[236]. Also, *Ric-8b* was shown to enhance the ability of $G\alpha_{olf}$ to induce cAMP production in heterologous systems[237] (Figure 1.15). RTP1S, along with RIC8B and the Rho tag act synergistically to maximise the luciferase responses in HEK293 cells[236].

Several combinations of all these different factors have been employed to deorphanise several receptors[215, 221, 222, 224, 229, 231, 238], alone or in combination with some of the other *in vitro* or *in vivo* techniques mentioned above. Nearly all have used the addition of a tag, which might affect the binding specificity of some receptors. Recently, a cleavable signal peptide was shown to promote surface expression of ORs, in combination with the trafficking proteins. Given that the tag is cleaved, the final receptor protein is virtually intact and should provide more reliable responses[239]. Another variable between studies is the $G\alpha$ subunit used; initially, generic $G\alpha_{15}$ or $G\alpha_{16}$ subunits were used, whereas later, $G\alpha_{olf}$ became more popular. Comparison of the responses obtained

by using either of these, demonstrated that the odorant response profile can be modified depending on which subunit is used[240]. Therefore, adhering to $G\alpha_{olf}$, the subunit present in OSNs, is preferable.

Higher rates of surface expression of OR proteins has made possible some large-scale deorphanisation efforts. Saito et al.[238] cloned a couple hundred mouse and human OR genes that cover the different subfamilies of the repertoire, and screened them against a panel of 93 odorants that represent varying functional groups, sizes and structures. Specific and reproducible responses could be obtained for a quarter of the screened mouse ORs and 4% of the human receptors. Taken together, the interaction matrix confirmed previous observations; different ORs show varying breath of tuning, with some responding to many odorants with dissimilar molecular structures, while others respond only to closely related compounds. Stimulation with enantiomers resulted in differential activation patterns. The agonists for class I versus class II receptors only differed in that the former tend to be more hydrophilic, which is consistent with their evolutionary origin in fish[238]. A similar study later on revealed an additional 27 receptor-ligand interactions for human ORs[241]. High-throughput studies like this one, allow a more comprehensive characterisation of the combinatorial code of olfactory coding, and further understanding of the rules governing the receptor-odorant interactions. The main caveat is that, for some odorants, the presence of mucus and an air-based delivery system might change considerably the activation profile[242, 243].

1.3.3 Antagonism.

Antagonists are odorants able to bind to ORs without activating the signal transduction pathway. As such, they compete for binding with agonists and can block their response. An example for the mOR-EG receptor was identified; when stimulation was performed with both eugenol and either methyl isoeugenol, isosafrole or oxidatively dimerised isoeugenol (a compound found in isoeugenol that has been stored for prolonged periods), no activation of the receptor could be recorded. The suppression of the response was due to competitive binding, since presentation of higher concentrations of eugenol versus the antagonists restored the activity.

Given that both agonists and antagonists bind the same receptor, they tend to have related structures and molecular properties. Exploiting this reasoning, Peterlin et al.[244] screened the I7 receptor with a set of compounds related to octanal and identified those that failed to elicit a response. When these were tested in conjunction with octanal, many were able to abolish the activation of the receptor, demonstrating their

antagonistic nature. This is particularly important, given that many screening efforts have grouped structurally related odorants into mixtures that are then applied to the receptors; only when a mixture elicits a response, the individual odorants are tested separately. The presence of both an agonist and antagonist might be the source of a high false negative rate in this strategy[230].

1.3.4 Adaptation and desensitisation of olfactory sensory neurones.

The sense of smell encounters complex mixtures of odorants. Perception is dependent on the different molecules involved, their particular concentrations and the relationships between them and the available receptors for activation. Additionally, OSNs respond to the same stimuli differently depending on previous experience, due to a process called odour adaptation. This mechanism allows the modulation of olfactory responses to ensure maximal sensitivity is achieved across time, and prevents the saturation of the system so that different odorants can always be detected. Adaptation is observed as a decrease in the elicited response by an odorant when it is presented repeatedly or when maintained as a constant stimulant. The process is reversible and normal activity levels are regained after the stimulation ceases. Adaptation is achieved by removing the OR proteins from the plasma membrane, by downregulating the expression of the OR gene and by reducing the activity of different signalling components in the transduction pathway[245].

Calcium influx through the CNG channels is a key step in the signalling pathway, both by its role in generating the action potential and as a regulator to allow odour adaptation. For the latter, it often couples to the calcium-binding protein calmodulin (CaM). The mouse CNG channel is composed of three different subunits: CNGA2, CNGA4 and CNGA1b. In the resting OSN the cytoplasmic concentration of Ca^{2+} is low, and CaM binds the CNGA4 and CNGA1b subunits in its Ca^{2+}-free form, also referred to as apocalmodulin. These two subunits have CaM binding sites for the interaction. When an OSN is activated and Ca^{2+} ions enter the cell, rapid association with apocalmodulin lowers the affinity of the CNG channel for cAMP, leading to a shift back to its closed state[246] (Figure 1.16). Both CNGA4 and CNGA1b CaM binding sites are necessary for the rapid modulation by Ca^{2+}-CaM[247]. The desensitisation of the CNG channel by Ca^{2+}-CaM is necessary to achieve a rapid termination of the response in OSNs; if the Ca^{2+}-CaM modulation is abrogated by mutating the binding site in CNGA1b, OSNs

have much slower decay rates after stimulation[248].

Ca^{2+}-CaM is also able to stimulate PDE1C, a phosphodiesterase highly expressed in OSNs. Upon the calcium rise and its association with CaM, PDE1C hydrolyses cAMP several fold more efficiently, also contributing to the closure of the CNG channels[249]. Furthermore, Ca^{2+}-CaM activates CaM kinase II (CaMKII), which is able to phosphorylate ACIII and stop the generation of cAMP[250] (Figure 1.16). In OSNs treated with CaMKII inhibitors, the onset of adaptation is reduced and the recovery occurs faster[251]. All these different processes are activated with the calcium rise, to block the transduction pathway at different levels and inhibit further action potentials from occurring to the same stimuli.

The inhibition of responses to repeated or continuous stimuli also includes targeting of the OR proteins directly. Odorant stimulation provokes the localisation of the GPCR kinase 3 (GRK3, also known as βARK2)[252] and the cAMP-dependent protein kinase A (PKA)[253] to the plasma membrane. In here, they associate with the ORs and phosphorylate them[253], which in turn makes them targets for β-arrestin2[252, 253] (Figure 1.16). Blockage of GRK3 results in higher levels of cAMP upon odorant stimulation[252] and also avoids the decline in the activation response, considerably slowing down the termination kinetics[254]. The activated receptors, targeted by β-arrestin2 are then engulfed by clathrin-coated vesicles and internalised, impeding them from further inter-

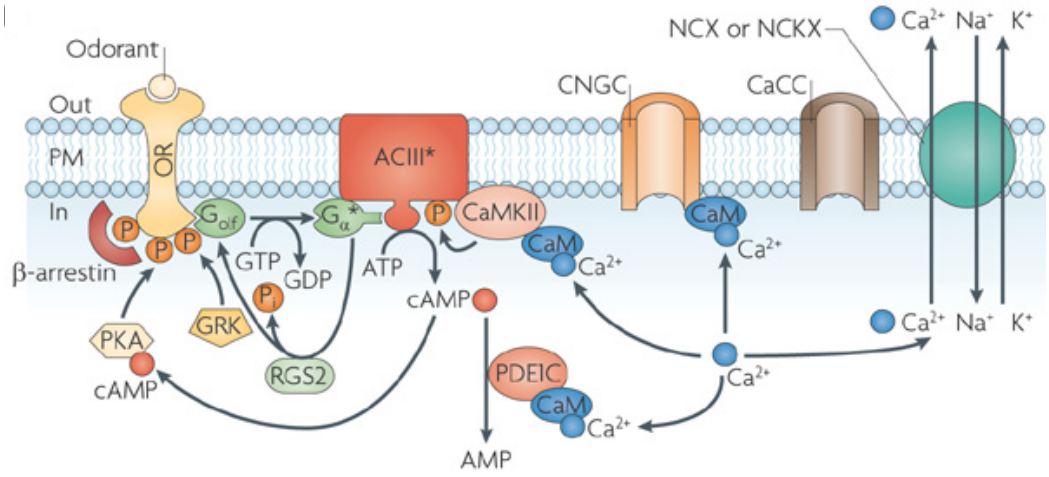


Figure 1.16 – Adaptation of OSNs to repeated stimulation. Upon stimulation of an OSN, Ca^{2+} influx induces adaptation by regulating the activity of key players in the signal transduction pathway. Ca^{2+} bound to calmodulin (CaM) interacts with the CNG channel and diminishes its affinity for cAMP to shut it down. Also, it interacts with phosphodiesterase 1C (PDE1C) to increase the hydrolysis of cAMP. In conjunction with the Ca^{2+}-calmodulin-dependent kinase type II (CaMKII), the activity of ACIII is diminished. Additionally, the OR proteins are phosphorylated by protein kinase A (PKA) and the G protein-coupled receptor kinase (GRK) which makes them targets of β-arrestin. These are then internalised to avoid further interactions with the ligands. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Neuroscience ([47]), copyright (2010).

action with ligands. This trafficking is strictly dependent on the proper phosphorylation of the receptor proteins, which is necessary for β-arrestin2 binding[253].

1.3.5 From detection to perception: impact of functional variation.

The strategy employed by the olfactory system to recognise millions of different odorants is based on the creation of a large and diverse set of receptors that can accommodate the structural diversity of ligands into a combinatorial code. The expansion of the OR repertoire has occurred through constant duplication and diversification events[54–57, 61]. Copy number variants (CNVs) are a form of structural variation usually defined as the deletion or duplication of a genomic segment greater than 1 kb. Not surprisingly, the catalogue of CNVs among human genomes is enriched for OR genes[255]. Several groups have analysed the CNVs affecting ORs in diverse sets of individuals, ranging from only a couple dozen to several hundred. On average, one third of the OR repertoire is affected by CNVs, without distinction between genes and pseudogenes[255, 256]; however, if only deletions are analysed, there is a greater number of pseudogenes affected (9.5 pseudogenes versus 3.8 genes per individual) and this is even stronger for homozygous deletions (3.9 versus 0.4 per person)[257]. The variation events tend to cluster in hotspots, which suggests that large deletion and duplication events, including several receptor genes, are common[256, 257]. Analysis of 150 genomes from the 1000 Genomes Project (1000GP) data revealed 313 copy-number variable loci involving OR genes. These included similar numbers of deletions and amplifications, which ranged from 3 up to 9 copies for a given OR[257]. Whereas the effect of additional copies of a particular receptor is not obvious, deletions are likely to influence the detection of their preferred ligand, at least for those odorants detected by a relatively small number of ORs. In all studies, a quarter to over half of the individuals analysed harboured at least one homozygous deletion and, in some, there were as many as four[256, 257].

The evolution of the OR genomic architecture is an ongoing process. As such, a considerable proportion of receptors with deleterious mutations are found segregating between their intact and pseudogene forms in the population. Some pseudogenes have only one disruption in their ORF, suggesting they are recent events. When these were genotyped in several people, up to half were found to be indeed segregating pseudogenes (SPGs)[258, 259]. What's more, when all SNPs, small indels and structural variants were taken into account, 59% of the receptors annotated as intact were SPGs. On average

each individual had 35 interrupted alleles, 11 of which were homozygous disruptions. Overall, the data analysed indicated that every person has a different combination of intact and pseudogenised alleles[260].

One such SPG has been directly linked to a specific anosmia to the compound isovaleric acid (IVA). Menashe et al.[261] tested the detection threshold of four odorants in 377 individuals; additionally, they genotyped the participants for SNPs that result in loss of function alleles for 52 OR genes. A strong association was evident between a SNP in *OR11H7P* and the sensitivity to IVA. Individuals carrying two copies of the pseudogene form of this receptor were only able to detect it at high concentrations. Consistently, *in vitro* experiments confirmed the interaction between the intact form of OR11H7P and IVA at a wide range of concentrations; also, the two neighbouring receptors responded to IVA at the highest concentration tested, but not at lower thresholds[261]. The insensitivity to IVA had already been observed, thirty years before, in a different species. Two C57BL strains of mice were shown to be anosmic to IVA whereas many other laboratory strains could readily detect it[262]; and the phenotype was later on linked to a locus in chromosome 4[263], which contains a cluster of OR genes[55].

As mentioned previously, a single amino acid change can shift the receptor's affinity for different odorants[224]; further accumulation of additional mutations eventually results in the ability to bind a different set of ligands[87]. This phenomenon is evidenced when the agonist profile of orthologous ORs in different species is compared to that of the paralogous receptors in a given species. A comparative study selected deorphanised human ORs and identified the orthologous receptors in the chimpanzee and macaque. Similarly, deorphanised mouse ORs were compared to their rat counterparts. When tested in a heterologous system, the orthologous receptors responded to the same set of agonists 82% of the time. In contrast, the paralogous human ORs, pertaining to the same subfamily, responded to the same ligand only 33% of the time. Interestingly, orthologs consistently showed differences in affinity, with some receptors being more sensitive to the same ligand in particular species[264], in line with what was observed for the mouse and rat I7 receptor in response to octanal[224]. Thus, differences in perception are not accounted by loss of function mutations alone, but are also influenced by variation that modifies the sensitivity of the functional receptors.

SNPs and small indels are crucial to generate diversity in the receptor repertoire, not only to allow divergence of paralogs, but also to create variable alleles of the same receptor. Mainland and colleagues mined the 1000GP data and found that the ORs annotated as functional receptors in the human reference genome had a median of five

different alleles at a frequency of 1% or greater. In fact, less than 5% had a unique allele. What's more interesting, is that when they tested 46 different alleles for 16 OR genes, 63% of the receptors with polymorphisms had differences in their *in vitro* responses to agonists. When this was compared to the 1000GP data, it indicated that any two individuals carry alleles that respond differently to a given ligand at over 30% of their OR intact repertoire[241]. Several examples exist of associations between particular ORs and SNPs that alter their affinity for their ligands; most of the work has been carried out in humans, because of the ease to ask participants to rate the intensity and/or valence of a given odorant. OT10G4, for example, binds with highest affinity guaiacol; a set of individuals were asked to rate the intensity of a solution of this odorant and their scores were correlated to the alleles they carried for this gene. There were four alleles with a minor allele frequency higher than 4%, and each was tested *in vitro*; two of the alleles had significantly lower affinity for guaicaol and, consistently, participants carrying these alleles rated the odorant less intense and more pleasant[241].

The advent of genome wide association studies (GWAS) has allowed the identification of links between genomic variation and many phenotypic traits and diseases. Much earlier studies had already suggested that the ability to detect certain odorants was heritable. For example, inspection of the pedigrees of 36 families bearing individuals that were unable to detect the musk pentadecalactone revealed that this trait was inherited as a simple recessive autosomal character[265]. Thus, association studies have been useful to identify OR variation that influences perception of certain odorants. In some remarkable cases, a single variant has been able to explain almost all the variation in sensitivity to a particular odorant, whereas in other, the contribution of one receptor is relatively small. In the case of β-ionone, a nonsynonymous SNP in the *OR5A1* receptor can explain 96.3% of the variation in perception of this compound. The detection threshold in the population spans five orders of magnitude and is bimodally distributed; the sensitive allele is dominant[266]. In contrast, variants in *OR2J3* affect the affinity of this receptor for cis-3-hexen-1-ol (the smell of cut grass), and they explain only 26.4% of the phenotypic variance. There are several haplotypes in the population for this OR, each conferring a variable sensitivity threshold. In heterologous systems, it was shown that two of the SNPs reduced the affinity of the receptor and, when together, completely abolished binding; subjects carrying both mutations had significantly higher threholds of detection[267].

Another example of specific anosmia is that to the compound androstenone, which is produced by male pigs and can be found in pork of fertile males[268]. The ability to

smell this odorant is also heritable[269], and is influenced by the sequence of the *OR7D4* gene. Some individuals carry two nonsynonymous SNPs in linkage disequilibrium that abolish binding to androstenone *in vitro*[270]. Consequently, the people that carry one or two of these insensitive alleles, were much more less able to detect the odorant, and they rated it more positively. Whereas the sensitive-allele carriers described the smell as urine-like or sweaty, people with insensitive alleles would refer to it as sweet and vanilla-like[268, 270]. What's more, it could be demonstrated that the genotype of participants tasting meat samples with varying concentrations of androstenone influenced how much they disliked the products[268]. This was also observed for β-ionone, when added to products as different as chocolate, fruit juice or household fragrances. There was a clear correlation between the sensitivity of their receptor alleles and their likeness of the products, demonstrating a direct influence on perception and consumer preferences[266].

Overall, it is clear that the combination of active and inactive receptors each of us carry influences our abilities to detect different odorants. Furthermore, the combinations of functional alleles determine the sensitivity to many other compounds and influence our perception and behavioural responses to them. The great variance in the composition of each person's receptor repertoire results in an individualised sensory experience and implies that each nose smells the world differently[241, 260, 271] (Figure 1.17).

Figure 1.17 – Individualised OR repertoire leads to unique perception. Around 30% of the OR repertoire is different between any two individuals. The different ORs expressed in each nose are represented as coloured shapes. Each individual shares different receptors with different people and some individuals lack particular receptors. Some receptors shared by different people have variation that can alter their affinity to their ligands (such as the solid shapes, compared to the ones filled in white). The particular combination of ORs in each person's nose lead to a unique perception of the olfactory world. Reprinted by permission from Macmillan Publishers Ltd: Nature Neuroscience ([272]), copyright (2014).

1.4 Plasticity of the olfactory system.

The olfactory system is constantly surveying the environment and conveys a lot of information about it. Animals are capable of associating olfactory cues with different objects and situations, and every encounter is an opportunity for learning what all those odorants mean. The constant addition of newborn neurones to both the MOE and the MOB, represents an opportunity to shape and refine the set of receptors and their associated circuits in the brain; this allows maximisation of information coding and adaptation to the particular niche the animal lives in. Adult neurogenesis is well characterised in mammals and occurs only in two brain areas: the subgranular zone of the dentate gyrus in the hippocampus and the subventricular zone (SVZ). Adult neurogenesis slows down as animals age. Neuroblast progenitors generated in the SVZ migrate along the rostral migratory stream (RMS) until they reach the olfactory bulb. Once in there, they migrate radially to integrate into the different MOB layers, where they mature into inhibitory interneurones, specifically granule and periglomerular cells. These inhibitory interneurones wire to mitral and tufted cells to regulate the processing of olfactory information. However, only around half of the produced neurones survive for more than a month[33, 273], which might reflect the difficulty to integrate into an already developed network.

The correct integration and survival of newborn neurones are influenced by activity-dependent mechanisms. Elimination of sensory inputs by unilateral naris occlusion (UNO) resulted in a decrease in the number of newborn interneurones that integrated into the the MOB. This effect was specific to the period when neurones were 14 to 28 days old. Deprivation before or after this had no effect, indicating that there exists a critical period when sensory activation is crucial for the survival of the new neurones[274]. In opposite experiments, animals were either constantly exposed to varied olfactory stimuli during 40 days[275, 276], or trained in an olfactory association learning task[277], and this resulted in an increased number of newborn neurones present in the MOB, measured by BrdU positive cells. Since the rate of neurogenesis was not affected, the difference was due to enhanced survival[275–277], which was also supported by an observed decrease in apoptotic cells[278].

Stimulation with odorants results in the expression of immediate early genes (IEGs), which are turned on upon electrophysiological activation. Expression of such genes could be detected in a quarter of newborn neurones after presentation of eight sets of three novel odorants each; this implies that these cells were responsive and had integrated

into the network that processes olfactory information. Responses to novel odorants, as measured by expression of IEGs, could be detected at least up to four months after the birth of the neurones. However, the percentage of responding cells decreased with increasing neurone age. Interestingly, if the animals were constantly stimulated, the number of activated newborn neurones remained high[279].

In control conditions, a high proportion of the newborn neurones that reach the MOB die after a few weeks[277]. However, when animals learn to associate an odorant with a reward, some of these new neurones are used to encode such information, and their survival depends on the ability to retain the information. Animals trained in this behavioural paradigm during five days, were able to remember the association after five days post-training, but not after 30 or 90 days. At five days, a high proportion of the newborn cells were activated upon recall of the association, suggesting that they were actively involved in learning the task. The number of surviving newborn neurones was higher at 5 days post-training, and remained high at 30 but drastically decreased at 90. Even though animals were not able to remember the task 30 days after training, when they were re-trained they learned it much faster, suggesting the task was partially preserved; this was not achieved at 90 days, when the neurones had already died. Further, treating the animals with a drug that enhances learning, resulted in retention of the task both at 30 and 90 days post-training and, consistently, the number of surviving neurones remained high[280].

Therefore, a model emerges in which newborn neurones arriving to the MOB are recruited to encode an olfactory learning task, but when the task is forgotten the neurones die. This was directly tested by *erasing* the olfactory memory by following the olfactory conditioning with a visual conditioning paradigm, while randomly presenting the olfactory cue. These animals were not able to remember the olfactory association and had a decreased survival rate of newborn neurones, compared to animals that were able to remember. What's more, when cell death was pharmacologically blocked, the olfactory memory was retained along with the newborn neurones involved[281].

Activity dependent stimulation therefore seems to enhance the survival of the newly arriving cells to the MOB. Many studies have corroborated that such enhanced survival correlates with better performance in olfactory learning and discrimination tasks. For example, Mandairon et al.[261] found three pairs of odorants that were similar to each other, such that mice failed to differentiate between them. Animals were exposed to one pair of these odorants for 20 days and tested again. The animals in the enriched environment were able to discriminate all three pairs of chemicals, even though they were

presented only one during training. The authors hypothesised that this was because the different odorant pairs activate overlapping parts of the MOB and, therefore, stimulation with one pair impacts the responses to all of them[282]. Similar results were obtained when instead of a passive presentation of the odorants, the animals were subjected to a forced-choice discrimination test[283].

The enhancement of olfactory learning and neurone survival by odour enrichment is dependent on the paradigm used. A study compared the effects of presenting a different odorant every day, versus using a mixture of all different odorants. The authors hypothesised that novelty was important to achieve the observed effects; indeed, the group that received the same complex mixture of odorants every day performed at the same levels as control animals, in a two-trial recognition test, while the group that was exposed to a different odour every day showed enhanced olfactory memory. Consistently, only the latter group had an increased proportion of newborn cells[284]. Another group used a similar reasoning to propose that social isolation results in olfactory monotony and showed that singly-housed mice are unable to recognise an individual that was presented 24 hours before, while group-housed animals had no trouble remembering it as familiar. To demonstrate that this lack of social long term memory was the result of a lack of olfactory stimulation, they enriched the environment of the singly-housed animals with either fruit essences or soiled bedding. The animals stimulated in this way performed as well as group-housed mice in social recognition tasks[285].

In the MOE there is also a constant turnover of the OSNs. Such a process opens a door to tailor the constituents of the overall neural population, to adapt to the environment and maximise the appropriate detection and downstream responses to the odorants that are encountered. Such an example was presented in 1993 by Wang and colleagues, that demonstrated that mice that are unable to detect androstenone or isovaleric acid at low concentrations, could enhance their sensitivity by repeated exposure to those compounds[86]. The mechanisms behind such sensitisation are now starting to be unraveled. It has been demonstrated that the stimulation of an OSN type with its cognate ligand retards its apoptotic cycle. For these experiments, mice were infected with an adenovirus carrying the OR *I7* and GFP, and then were exposed to octanal for six weeks. The animals that received octanal stimulation had a much larger number of GFP^+ cells compared to unstimulated controls, suggesting that these cells had survived longer. Further, this effect was specific to OSNs expressing I7[286]. This process was mediated by the expression of *Bcl2*, an anti-apoptotic factor, mediated by the MAPK/CREB and PI3K/Akt pathways upon odorant stimulation[286, 287].

Complementary experiments have induced generalised OSN apoptosis to determine what factors can stimulate survival even in such conditions. Removal of the olfactory bulb causes the death of all OSNs by depriving them from trophic support. Such an operation can be performed in one half of the bulb only, and the intact side serves as a control. Mice were exposed to several single and complex odorants for two days and then unilateral bulbectomy was performed; interestingly, a population of OSNs remained in place, while all neurones were lost in controls. To demonstrate that the neurones surviving were those that were stimulated by the presented odorants, the same experiment was done by infecting with the *I7* adenovirus and stimulating with octanal; as expected, the I7 expressing neurones survived after the operation, but only if the odorant presented was octanal[286].

A similar approach exploited the expression of endothelin in OSNs, another anti-apoptotic factor. Rat pups were treated with an antagonist for the endothelin receptor, thus blocking its effects and increasing the apoptosis of mature OSNs. These animals had reduced EOG recording responses to odorants; however, the pups were still able to detect and locate their mother odour, suggesting that the OSNs involved had survived. Indeed, when the pups were treated with octanal, many more I7-expressing neurones survived, once more indicting that active neurones have enhanced survival, even in conditions of induced apoptosis[288].

Another indication on the importance of activity for OSN survival came from the *Cnga2* KO animals. As explained earlier, this gene is an essential component of the CNG channel that allows calcium entry to the OSN upon odorant stimulation. Without a functional CNG channel, neurones are not able to fire action potentials. Since *Cnga2* is located in the X chromosome, heterozygous females contain a mosaic population of OSNs expressing either the functional or the knocked-out version of the gene, depending on X chromosome inactivation. In young animals this was indeed the case and both populations of neurones projected axons to the MOB. However, the OSNs lacking *Cnga2* were progressively depleted and adult animals contained only neurones expressing the functional *Cnga2* allele. Interestingly, when odorant stimulation was blocked by UNO both types of OSNs remained in the MOE. This suggests that in competitive conditions, odorant-evoked activity is necessary for the survival of OSNs[289].

On the other hand, UNO experiments on wild-type mice have shown that OSNs expressing different ORs are affected disparately by the deprivation of stimulation. 15 ORs were assessed by *in situ* hybridisation in both the occluded and open sides of the nose; half of the genes were found more frequently in the occluded side, but some were

also less represented and others did not change[290]. A microarray analysis also found that the general trend was for ORs to be upregulated by sensory deprivation, along with essential components of the signalling pathway[291]. A confounding effect in these studies, however, is that the occluded side not only is devoid of odorant stimulation, but also of pathogens and general insults that would, in normal conditions, have an impact on the shaping of the neuronal population.

A recent study provided insight into a molecular mechanism behind enhanced OSN survival by olfactory stimulation. Santoro and Dulac identified a histone variant, H2BE, that replaces the canonical version (H2B) specifically in the sensory neurones of the MOE and VNO. The expression levels of H2BE in each OSN varied widely and were correlated to the OR expressed by the neurone. By studying both a knock-out and an overexpression mouse model, the authors identified a correlation between the levels of the histone variant and the life span of the OSNs: those neurones that expressed low levels of H2BE lived for longer periods than the OSNs with high H2BE expression. What's more, the levels of H2BE were determined by the activity of the OSN. Therefore, those neurones that were constantly activated by their cognate ligand, had reduced H2BE and survived for longer in the MOE. Over time, this results in an enrichment of active OSNs and a depletion of inactive OSNs in the population. Such a mechanism may also explain the results observed in the UNO experiments; whether an OR goes up or down on the occluded side correlates with its initial levels of H2BE expression[292].

A couple of studies have shown contradicting evidence to the data presented above. Cavallin et al.[293] found that UNO decreased the number of OSNs expressing M72, but this was also the phenotype after exposing the animals to acetophenone (the ligand for M72) for a month. Similarly, exposure during three weeks to lyral, the ligand of MOR23, resulted in a decrease of 70% in the population of OSNs expressing such receptor[294]. More puzzling was the fact that the reduced number of cells that remained in the MOE of treated animals expressed higher levels of both the receptor mRNA and protein. Thus, the global levels of MOR23 across the whole MOE were not changed, but the number of OSNs was. These results, however, were specific for MOR23, since exposure of animals to acetophenone did not alter the number of cells expressing M71 or the levels of receptor mRNA per cell[294]. It is difficult to compare these results to the other studies, since each uses a different exposure paradigm; the concentration of the odorant, the time course of the experiment, how often and for how long the odorant is presented, etc. are all different between studies.

Nonetheless, a robust body of evidence supports a model where active OSNs have

enhanced survival, which leads to their enrichment in the overall neural population of the MOE. This could result in more efficient detection and processing of those odorants that are constantly encountered by the animal. These data suggest that the system is plastic enough to adapt to changes in the environment and the presence of novel odorants.

The present dissertation contains the results from the development of a technique, based on high-throughput RNA sequencing (RNAseq), to profile the expression of the complete receptor repertoire of the MOE and VNO of mice. I will first present evidence on the suitability of this technique to profile the transcriptome of the mouse olfactory system, and how it compares to other established technologies. Then, I will show the application of the technique to decompose the transcriptome of the MOE, from the whole tissue to single OSNs. Finally, I will present evidence on the mechanisms behind the regulation of the expression of the complete OR repertoire, based on data from wild-type mice of different gender, strain or exposed to different olfactory environments.

Chapter 2

The transcriptome of the mouse olfactory system.

Some of the material presented in this chapter has been previously published in reference [295]. I confirm the work is my own unless otherwise stated and I have ownership of the copyright for its reproduction.

Much is known about the structure and composition of the MOE and VNO, and of the essential components expressed by their sensory neurones to achieve olfactory detection and signalling. Much less understood is the transcriptional profile of these tissues and their individual components, how specialised they are and whether transcriptional differences can account for observed disparities in olfactory-mediated behaviours. On the study of the olfactory system, the sensory neurones are, undoubtedly, the most interesting component, given their role in signal detection and information processing. The regulation of the characteristic expression pattern of the receptors –monogenic and monoallelic– is still incompletely understood. Not surprisingly, a lot of work has been carried out to tackle the fundamental questions regarding receptor expression and regulation.

The sequencing of the mouse genome allowed the use of computational methods to search for the the complete repertoire of OR and VR genes, based on homology searches with the few dozen experimentally obtained sequences from rat and mouse sensory neurones[58, 65]. Recent analyses of the mouse genome have identified 1,130 OR genes with an intact open reading frame (ORF) and an additional 236 pseudogenes[296].

In the case of VR genes, there are 392 V1R genes (239 intact)[136] and 279 V2R genes (158 intact)[65]. If all these receptor genes are indeed true ORs and VRs, they should be expressed in the MOE and VNO respectively. Expression of several OR genes has been detected in other tissues, and some have been shown to play a role in functions other than olfaction, such as sperm chemotaxis[297, 298]. While some can have both olfactory and non-olfactory functions, others might be specifically expressed outside the olfactory system and should not be considered part of the receptor repertoire.

The study of the expression of receptor genes has been a daunting task for several reasons. First, there is a very large number of genes to analyse; this makes approaches such as *in situ* hybridisation slow and labour intensive. Secondly, both ORs and VRs tend to have close paralogous sequences; in some cases, two receptors can be nearly identical which makes their discrimination challenging. And third, each receptor is expressed in only a small subset of the cells from the whole population found in the tissue, which leads to very low expression levels when analysing total tissue RNA samples.

The expression of a few hundred receptor genes has been confirmed by *in situ* hybridisation assays. In an effort to produce a high-throughput approach, Firestein and colleagues developed custom microarrays containing probe sets for the majority of the annotated OR and VR sequences available at the time. For the ORs, the Celera mouse genome was used, one of the earliest versions available; since then, the OR repertoire has changed considerably and many receptors have been identified as artefacts of assembly errors. In the case of the VR genes, a more recent assembly of much better quality was used. For both OR and VR genes, probes were designed based on computationally predicted 3' untranslated regions (UTRs) for two reasons: the array is biased to detect this portion of the transcripts, and UTR sequences tend to be more variable compared to the CDS. However, no evidence of the accuracy of these predictions was provided[299, 300]. Despite all these drawbacks, the authors reported that 817 (62.55%) of the ORs represented in their array were enriched in the MOE compared to the VNO[299]; and 168 (54.5%) and 98 (32.1%) V1R and V2R genes respectively were expressed in the VNO[300]. These studies represent the highest number of receptors with validated expression in their cognate tissue.

Microarrays have the advantage of providing space for thousands of probes; this makes it possible to interrogate all the genes annotated in the genome at the same time in a given sample. The high-throughput character of microarrays made them a very popular technology and have been widely used in different species, tissues and experimental designs. Yet, they have several limitations. First, microarrays suffer from high

background levels of hybridisation, that is, signal is recorded irrespective of the presence of a transcript; this makes them less sensitive to accurate estimation of gene expression levels, especially for lowly expressed genes where it becomes difficult to differentiate true signal from background. Secondly, some probes are more efficient than others in their hybridisation efficiency. Third, probes for genes with close paralogs can cross-hybridise, also making their accurate detection problematic. And fourth, the information obtained with the array is limited to known genes and relies on the correct annotation of the genome; thus, any novel genes cannot be detected[301].

A different technology, the NanoString nCounter platform, allows the simultaneous quantification of expression levels for hundreds of genes. Even though it does not provide genome-wide coverage, the sensitivity of the assay is much greater, especially for lowly expressed genes that can not be detected accurately by microarrays. This technology is based on the design of two probes for each gene: a *capture* probe, that contains 35 to 50 nucleotides complementary to the mRNA of interest, coupled to an affinity tag such as biotin; and a *reporter* probe, also consisting of 35 to 50 nucleotides of the target mRNA sequence, plus a tag that contains a specific sequence of different fluorophore-labelled RNAs that make up a unique code for identification. The assay consists on hybridising the RNA sample of interest with the capture and reporter probes; the hybridised probes are then fixed to a solid surface coated with streptavidin and imaged to identify the code of each molecule. Finally, counts for each type of targeted mRNA are obtained[302]. Using this technology, Khan et al. were able to design specific probes for 592 OR genes, based on their CDS; for the rest of the repertoire, probes were not specific enough to differentiate closely related genes. From the tested ORs, 577 (97.5%) were detected reliably in wild-type mice. They further demonstrated the sensitivity and specificity of the system with two experiments: 1) RNA was obtained from OSNs that expressed a particular OR gene and, as expected, a high signal was obtained for that OR only; and 2) RNA was prepared from a mouse strain that has a deletion in a cluster of about a hundred OR genes in chromosome 9; for the deleted genes, expression levels were no different from background while the rest of the repertoire was correctly detected[186]. Despite the great sensitivity and specificity of this system, it is restricted to the study of those receptor genes that are divergent enough to be unambiguously detected by short probes.

Nowadays microarrays are being replaced by high-throughput RNA sequencing (RNA-seq), given the drop in the cost of sequencing and the advances in the development of both robust protocols for library preparation and bioinformatic analysis tools for pro-

cessing the data. RNAseq has several advantages for transcriptome profiling. It has virtually no noise since all sequencing reads are obtained from the RNA present in the sample. It is quantitative: the number of sequenced reads per gene is proportional to its expression level, and there is no saturation of the signal; increasing expression of a gene results in increasing numbers of sequencing reads. It does not depend on *a priori* knowledge of the genome sequence or annotated genes, allowing for the discovery of new genes. Also, this results in a comprehensive capture of the different splice isoforms for a given gene, some of which may not be known. The major limitation of RNAseq is that it is based on short reads; these can sometimes map to several regions of the genome, which makes them ambiguous since their true place of origin can not be determined[303]; the increase in length of sequencing reads (from 32 bp to over a hundred bp) has ameliorated this problem, in conjunction with the development of paired-end sequencing that also increases the amount of sequence obtained from each molecule. Nonetheless, some regions of the genome are still affected by a high proportion of ambiguous reads –also called *multireads*– especially those containing genes with close paralogs.

At the beginning of my PhD, I had access to RNAseq data from both the MOE and VNO of C57BL/6J adult mice. Given the clear advantages of RNAseq over other transcriptomic technologies, I conducted a series of validation experiments to assess the suitability of this method to accurately profile the expression of the olfactory system and, mainly, of the receptor genes.

2.1 Transcriptome profiling by RNAseq.

In order to study the complete profile of genes expressed in the MOE and VNO of wild-type adult B6 mice, these tissues were dissected from both male and female mice and RNA was extracted to construct libraries for RNAseq. All samples were sequenced at high depth to ensure capture of lowly expressed genes, such as the receptors. The VNO samples, composed of the sensory neuroepithelium, progenitor and non-neuronal supporting cells, underlying glandular and cavernous tissue and a blood vessel with blood[1], yielded a mean of 37.1 $\pm$ 3.6 million (SD) paired-end fragments per sample. For the MOE, dissections included the underlying lamina propria, glandular tissue, olfactory nerves and blood vessels with blood, in addition to the olfactory epithelium[304]; for clarity, I will refer to these as the whole olfactory mucosa (WOM). WOM samples yielded 46.4 $\pm$ 4.3 million (SD) paired-end fragments on average (Table B.1 in Appendix B). All sequencing fragments were mapped to the mouse reference genome with a splice-

aware algorithm; on average, 80.96±1.2% (SD) and 87.78±3.0% (SD) of the VNO and WOM data respectively mapped unambiguously (Table B.2 in Appendix B). To estimate expression levels, I counted the number of fragments that mapped to each gene model, as annotated in the Ensembl database. In order to be able to compare expression levels across different samples, I normalised these raw counts to account for the depth of sequencing. Detailed methods can be found in Appendix A.

I first assessed the variation in gene expression among the three biological replicate samples for each sex and tissue. All pairwise comparisons showed very high correlation levels (Spearman's rank correlation coefficient rho > 0.95, p-value $< 2.2e\text{-}16$), with only small sets of genes showing unusually variable values among replicates (Figure 2.1). I therefore averaged the normalised counts for each gene in each tissue.

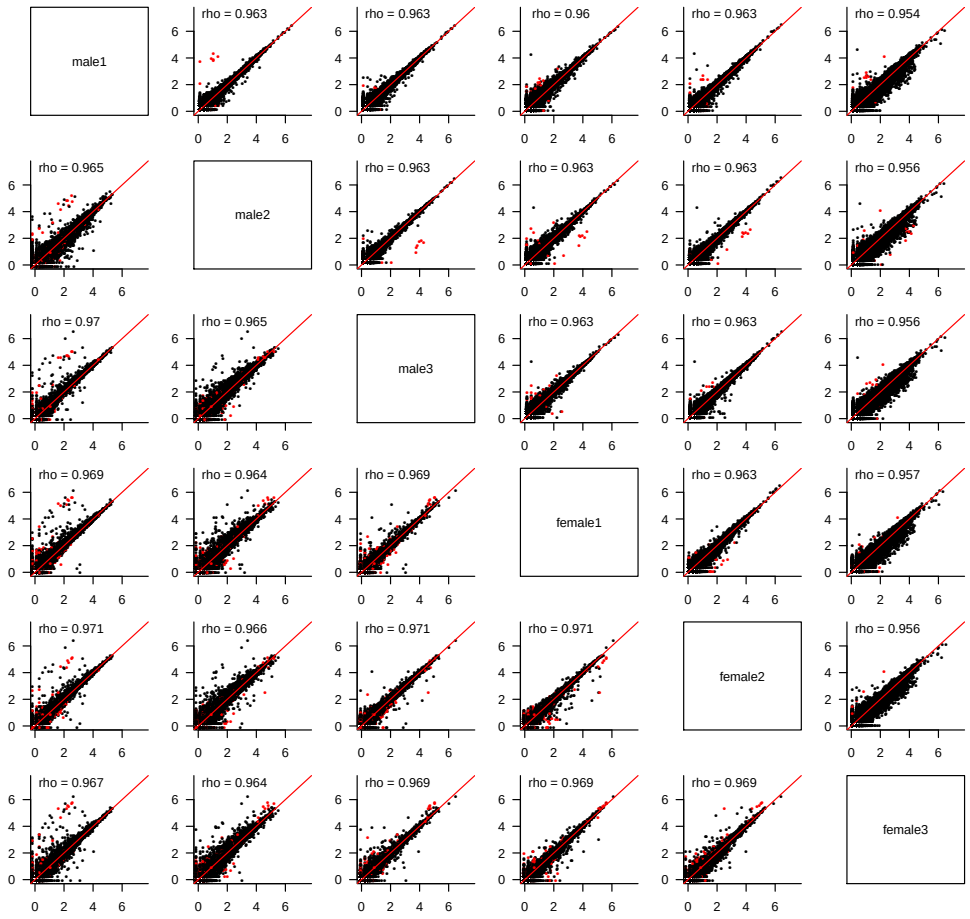


Figure 2.1 – Correlation between biological replicates. Scatter plots of the $\log_{10}$-transformed normalised counts for each pairwise comparison of biological replicates for the VNO (on top of the diagonal) and WOM (undeneath the diagonal) data. The 1:1 diagonal is in red. Samples are indicated in the diagonal. The Spearman's correlation value is on the top left corner. Genes with a coefficient of variation greater than the 90th percentile are highly variable between replicates, and are highlighted in red.

A total of 11,215 (29.01%) and 10,543 (27.28%) genes in the VNO and WOM respectively have no fragments mapped in any replicate, suggesting they are not expressed in that tissue. The expression of the remaining genes shows a bimodal distribution of low- and high-expressed genes characteristic of RNAseq datasets[305]. These can be decomposed into two normal-like overlapping distributions (Figure 2.2), and each gene can be assigned to either with a degree of confidence. Low-expressed genes typically do not have active chromatin marks, are enriched in non-functional mRNAs and, unlike the high-expressed genes, lack correlative protein expression data[305]. For the VNO and WOM respectively, 56.5% and 52.65% of the expressed genes are in the distribution of high-expressed genes with 50% probability or more.

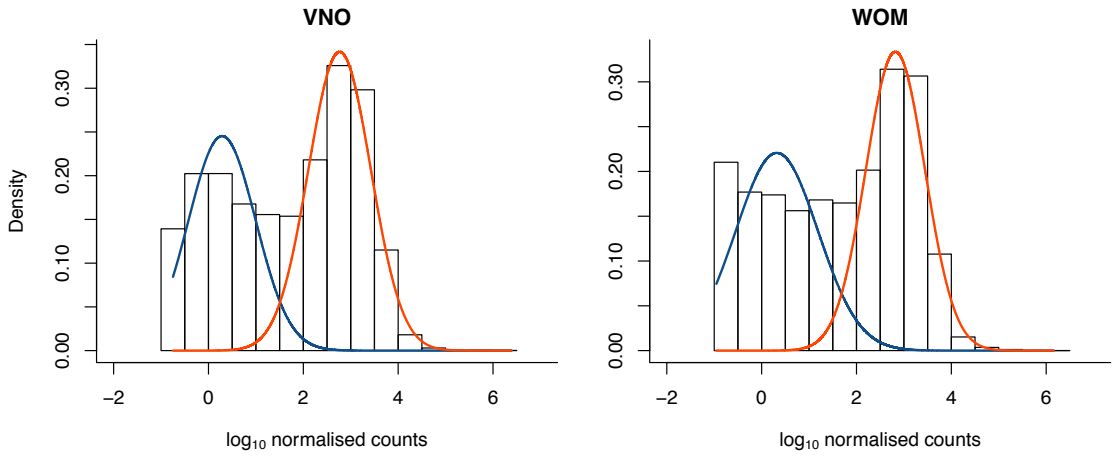


Figure 2.2 – The transcriptome shows a bimodal distribution of low- and high-expressed genes. Histograms showing the distribution of the mean $\log_{10}$-transformed normalised counts for all those genes expressed in at least one replicate, for the VNO and WOM. Superimposed are two normal-like distributions calculated from the data, that decompose the total bimodal distribution. The blue distribution contains all those genes expressed at low levels, while the red distribution represents highly expressed genes.

A comparison of the transcriptomes of both tissues revealed that 61.43% of the genes are differentially expressed (DE) with a false discovery rate (FDR) of less than 5%. To explore these further, I selected the genes with a fold change of four or higher and performed a functional terms enrichment analysis. As expected, those expressed higher in the VNO were enriched for VR genes, which contribute to the response to pheromones and detection of chemical stimulus involved in sensory perception. Additionally, the calcium signalling pathway, ionic and voltage-gated channel activity and regulation of vasoconstriction were significantly enriched as well. For the WOM, enriched genes were dominated by ORs and those involved in the olfactory transduction pathway, cAMP-mediated signalling and sensory perception. In addition, there was enrichment of ionic and ligand-gated channels and regulation of neurone, oligodendrocyte and glia differen-

tiation. In contrast, 'housekeeping' genes were expressed at similar levels in both tissues (Figure 2.3). All together, these data indicate that RNAseq is a highly reproducible technique to study the olfactory system, despite the heterogeneity in the composition of each tissue; and that the quantification of expression levels accurately reflects the principal functions of each tissue.

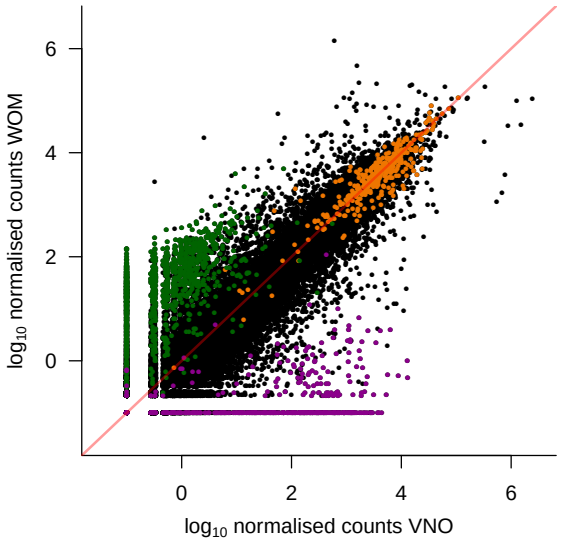


Figure 2.3 – The transcriptome of the VNO and the WOM. Scatter plot of the expression values for each gene in the VNO versus the WOM; 0.1 was added to the counts to be able to visualise those genes that are expressed in one tissue only. The VR genes (purple) are largely undetected in the WOM but clearly expressed in the VNO. The opposite is true for the OR genes (green). In contrast, housekeeping genes (orange) tend to be expressed at similar levels in both tissues. The red line represents the 1:1 diagonal.

2.1.1 Comparison to alternative methodologies.

As mentioned previously, the most popular technology for transcriptome profiling before the advent of high-throughput sequencing was expression microarrays. To assess whether RNAseq is a better technology to study the olfactory system, I used commercial Illumina microarrays to profile six more biological replicate VNO and WOM samples[1]. For both tissues, the overall expression values are correlated (rho = 0.66 for the VNO, 0.67 for the WOM, p-value < 2.2e-16). However, the RNAseq expression estimates are considerably different for genes that are either very lowly or very highly expressed. As can be observed in Figure 2.4A, the microarray intensity values flatten for the genes in both ends of the distribution, because of high background noise and saturation of the fluorescent signal; yet, in the RNAseq data, expression values span several orders of magnitude. Thus, the

[1]Microarray data are available in the ArrayExpress database under accession number E-MTAB-2163.

dynamic range of expression levels that can be detected is much broader for RNAseq compared to microarrays.

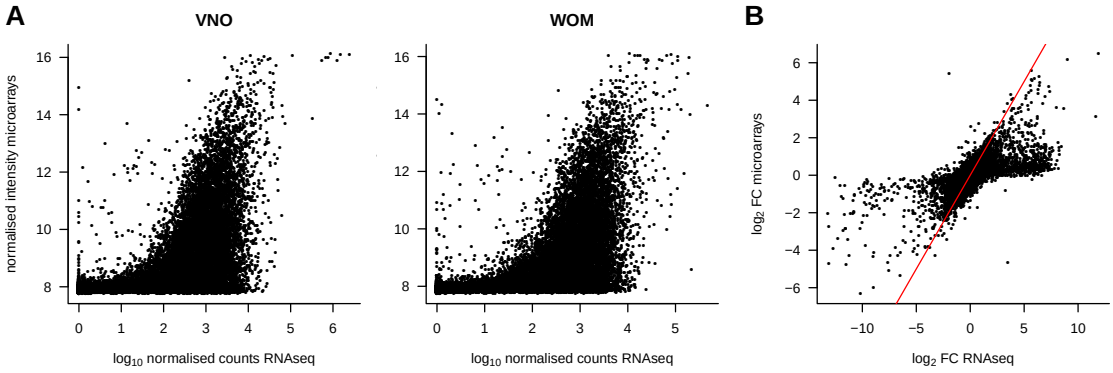


Figure 2.4 – Comparison of the RNAseq expression values versus the microarray intensity data. A) Scatter plot of the expression estimates for each gene from the RNAseq data compared to the microarrays. **B)** Comparison of the fold-change (FC) for each gene as assessed by both technologies. The red line indicates the 1:1 diagonal.

From the genes with expression signals above background in the microarray data, 11,499 can be directly mapped to an Ensembl gene id present in the RNAseq data. From these, 66.96% are reported as DE between the VNO and WOM samples (FDR $< 5\%$), a proportion similar to the results from the RNAseq data. 76% of the DE genes as assessed by the microarray are also DE from the sequencing data and the overall estimated fold-changes from both platforms are correlated (rho = 0.79, p-value $<$ 2.2e-16). However, a subset of the genes with large fold-changes identified as DE by RNAseq have small differences in the microarrays (Figure 2.4B). Previous studies have shown that the genes that are called as DE by only one technology validate well for the RNAseq calls but poorly for the microarrays[301], which suggests that these robust changes detected in the RNAseq data are likely to be true positives.

Quantitative real time PCR (qRT-PCR) is accepted as the gold standard for expression profiling, so next I compared both the RNAseq and the microarray expression estimates to a panel of qRT-PCR TaqMan gene expression assays[2]. I included data for genes with and without a known function in olfactory and vomeronasal signalling that cover the whole range of expression values observed (Figure 2.5). The correlation is considerably higher with the RNAseq data (rho = 0.87 for the VNO and 0.92 for the WOM, p-value $<$ 2.2e-16) than with the microarray estimates (rho = 0.70 for the VNO and 0.78 for the WOM, p-value $<$ 2.2e-16), indicating that RNAseq is better suited for transcriptome profiling in the olfactory system. As observed previously, these comparisons also

[2]TaqMan expression estimates were obtained and kindly provided by Maria Levitin.

suggest that microarrays are not sensitive enough to accurately detect genes expressed at low levels (Figure 2.5). Thus, the strong correlation between the qRT-PCR and the RNAseq data, strongly suggests that the expression estimates generated by sequencing are reproducible and specific, and provide a comprehensive characterisation of olfactory transcriptomes at a wide range of expression values.

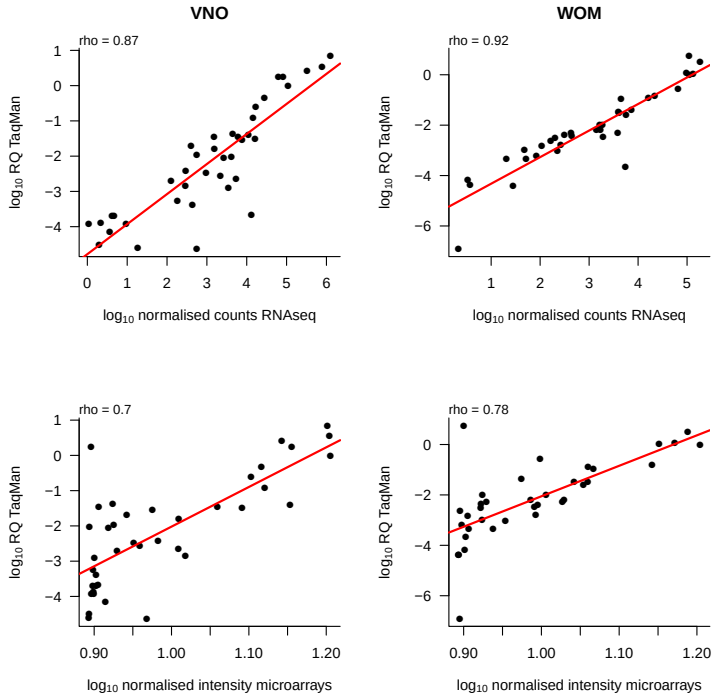


Figure 2.5 – Comparison with qRT-PCR TaqMan expression assays. The normalised RNAseq (top) or microarray (bottom) data was compared to expression estimates obtained by qRT-PCR TaqMan probes. The Spearman's correlation value is indicated in the top left corner. In red is the regression line. *B2m*, *Bpifa1*, *Calm1*, *Casp3*, *Clca1*, *Cnga2*, *Dmbt1*, *Dnase1*, *Eef1a1*, *Expi*, *Gnai2*, *Gnal*, *Gnao1*, *Gucy2e*, *H2-Aa*, *Ltf*, *Nelf*, *Olfr1213*, *Olfr123*, *Olfr124*, *Olfr1262*, *Olfr1347*, *Olfr1507*, *Olfr1509*, *Olfr222*, *Olfr691*, *Olfr692*, *Omp*, *Slc8a1*, *Stk32a*, *Trpc2*, *Vmn1r90*, *Vmn2r1*, *Vmn2r29*, *Vmn2r3* and *Xist* were assayed in both tissues and *Vmn2r121*, *Vmn2r28* and *Vmn2r6* in VNO only since amplification failed with WOM samples.

2.2 Expression of the receptor repertoire.

So far I have shown that the RNAseq data is reproducible and accurate at different expression levels. However, properly assaying the receptor repertoire has additional challenges. The monogenic expression of receptors in sensory neurones means each individual receptor is expressed in only a small subset of cells. Therefore the expression of any given receptor within the whole epithelium is low. The GRCm38 mouse assembly contains 1,250 annotated OR genes and 530 VR genes. To ensure that these represent the

complete repertoires, I took the cDNA sequences for the mouse VR genes as previously reported[65, 136], and aligned them to the genome with BLAST. I recovered 32 Ensembl genes that were not annotated as a VR gene, but that perfectly matched a VR cDNA sequence. These were included in subsequent analyses (Table B.3 in Appendix B). An additional 19 VR and 4 OR genes matched genes not annotated as receptors with high identity. These are annotated as novel genes and most likely represent additional members of the receptor repertoire. Proper analysis and annotation with new gene names is necessary for all these, which is beyond the scope of this dissertation; therefore, these were not considered in further analyses. It is worth noting that the above strategy was only intended to recover genes that are present in Ensembl but not labelled as VR or OR genes. Other methodologies such as hidden markov models would be more appropriate to discover new genes that possess the structure, domains and motifs that characterise these multigene families.

I first analysed the overall expression distribution for each class of receptors in their cognate tissue. In both cases, after normalising for gene length, the receptors in the repertoire do not have equal abundances, as may be expected if each receptor had equal probability of being chosen for expression by the OSNs. Instead we observe a large dynamic range of expression: a few receptors are expressed at high levels and the vast majority of the repertoire is expressed at relatively low levels. For the VR genes, the most highly expressed receptor, *Vmn2r89*, has a value of 4,363.4 normalised counts and only 22 other receptors have more than 1,000 normalised counts. In contrast, the median expression is 20.69 and 37.95 for V1R and V2R genes respectively (Figure 2.6). 436 VR genes (77.3%) have at least one fragment mapped uniquely. From the remaining 127, 70.1% are annotated pseudogenes, and 62 (48.8%) have multireads, which indicates that at least a fraction of these are expressed. 65 VR genes have no mapped fragments, either unique or multi-mapped, but these are all annotated as pseudogenes.

In the case of the OR genes the most abundant, *Olfr1507*, is expressed at 2,456.1 normalised counts and only 12 other receptor genes are above 500 normalised counts. The median expression is 26.88 normalised counts (Figure 2.7). Despite their relatively low abundance, 1,182 (94.56%) of all the annotated OR genes have at least one fragment mapped to their exonic region. Of the remaining 68 genes, 50 (73.5%) are annotated as pseudogenes and 17 have multireads that could indicate expression. Only 9 putatively functional OR genes have no evidence of expression in the WOM whatsoever (*Olfr115*, *Olfr141*, *Olfr504*, *Olfr564*, *Olfr574*, *Olfr834*, *Olfr1053*, *Olfr1061*, *Olfr1367*).

Importantly, the expression estimates for both OR and VR genes are consistent

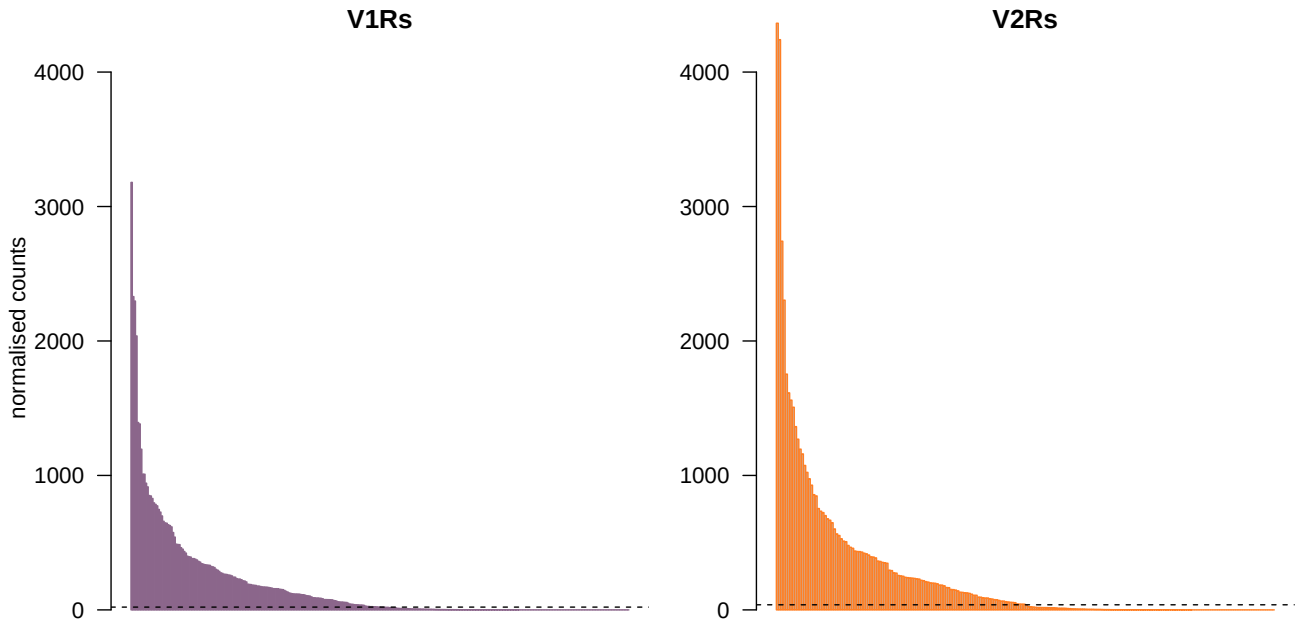


Figure 2.6 – **Expression of the VR repertoire.** Each bar represents the mean normalised expression of a V1R (purple) or V2R (orange) gene; counts have been normalised for gene length as well as depth of sequencing. The horizontal dotted line represents the median expression.

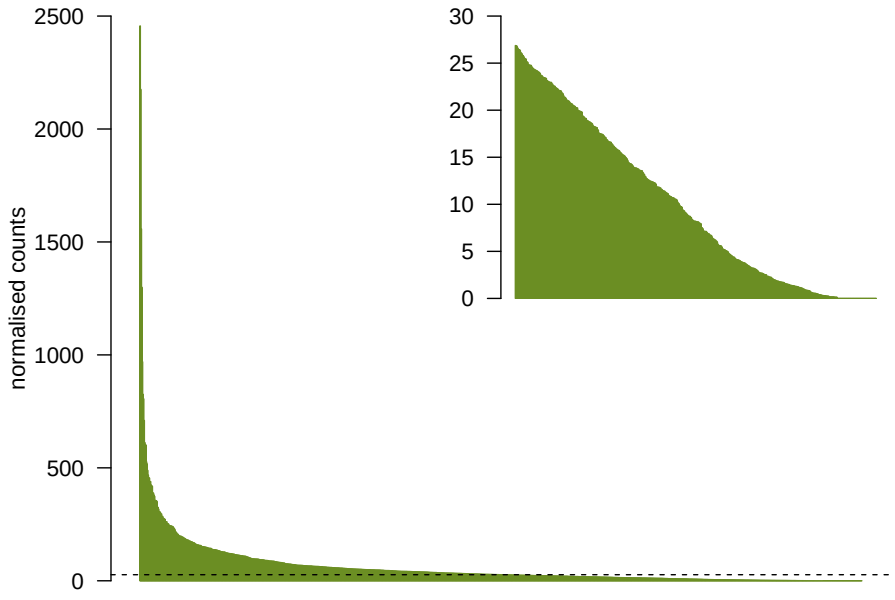


Figure 2.7 – **Expression of the OR repertoire.** Each bar represents the mean normalised expression of an OR gene; counts have been normalised for gene length as well as depth of sequencing. The horizontal dotted line represents the median expression. The inset contains all those receptors with expression values below the median.

between biological replicates. The median coefficient of variation (CV; standard deviation / mean) is 0.27 and 0.31 for the VR and OR genes respectively, suggesting that the uneven distribution observed is stereotypical. I could not identify any apparent pattern in the expression of either VR or OR genes based on cluster or genomic location. A difference in expression abundance is evident for genes and pseudogenes, with the former being expressed at much higher levels than the latter, for both VR and OR genes (p-value < 2.2e-16, Mann-Whitney test; Figure 2.8A-B). This is consistent with previous reports that observed the expression of pseudogenes was extinguished progressively after an OSN chose a functional receptor[207]. Interestingly, class I OR genes and V1R genes are also expressed at significantly lower levels than class II OR genes and V2R genes, respectively (p-value = 0.0077 for VRs and 0.0004 for ORs, Mann-Whitney test; Figure 2.8C-D).

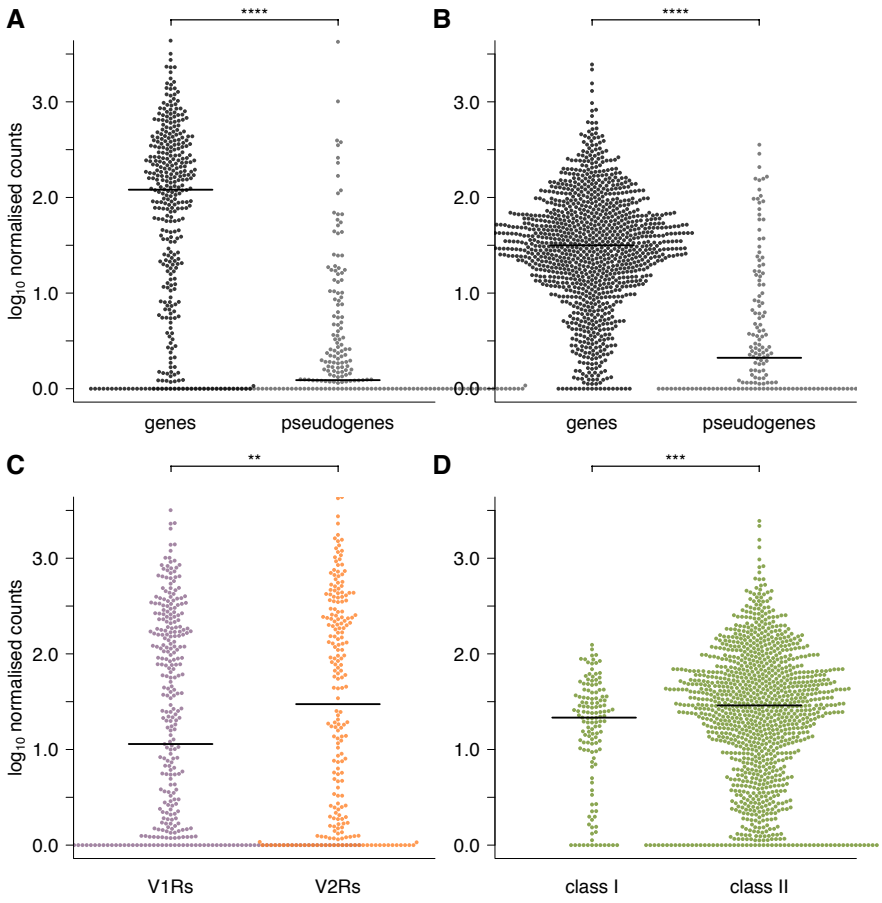


Figure 2.8 – Genes are expressed higher than pseudogenes. A-B) The normalised expression of each VR (A) and OR (B) gene is plotted, separated by whether they are annotated as intact genes or pseudogenes. Pseudogenes are expressed at much lower levels for both receptor types. **C-D)** Same but now separated by receptor class. V1R genes, as a whole, are expressed lower than V2R genes, and class II OR genes are expressed higher than class I genes. Mann-Whitney test; ** < 0.01, *** < 0.001 and **** < 2.2e-16.

Finally, I asked whether any VR genes are expressed in the WOM, or if OR genes are found in the VNO. One VR gene, *Vmn2r29*, is expressed in the WOM at a level that is higher than the median OR gene expression (32.1 normalised counts). This is consistent across all six replicates, suggesting there may be additional populations of sensory neurones in the MOE expressing receptors different to ORs. In the case of the VNO, 11 OR genes are expressed higher than the VR gene median, with *Olfr124* as the highest (496.7 normalised counts) followed by *Olfr692* and *Olfr1509* (262.2 and 126.1 normalised counts respectively). Both *Olfr124* and *Olfr692* consistently display higher expression values in the VNO than in the WOM.

2.2.1 Comparison to other methodologies.

In the commercial microarrays, there are probe sets for 178 VR genes and 1,051 OR genes, that can be mapped directly to a gene annotated in the Ensembl database; these represent 31.6% and 84.1% of the repertoires, respectively. The correlation between both datasets is only moderate for the OR receptors (rho = 0.54, p-value < 2.2e-16) and weak for the VR genes (rho = 0.30, p-value < 2.2e-16). As observed in Figure 2.9A, a proportion of the receptors accumulate in the lower detection threshold of the microarray intensity signal, close to the background level of the array; those genes show varied levels of expression in the RNAseq data, spanning several orders of magnitude. Therefore, the poor correlation can be attributed to the low sensitivity of the microarray to estimate expression of lowly expressed receptor genes.

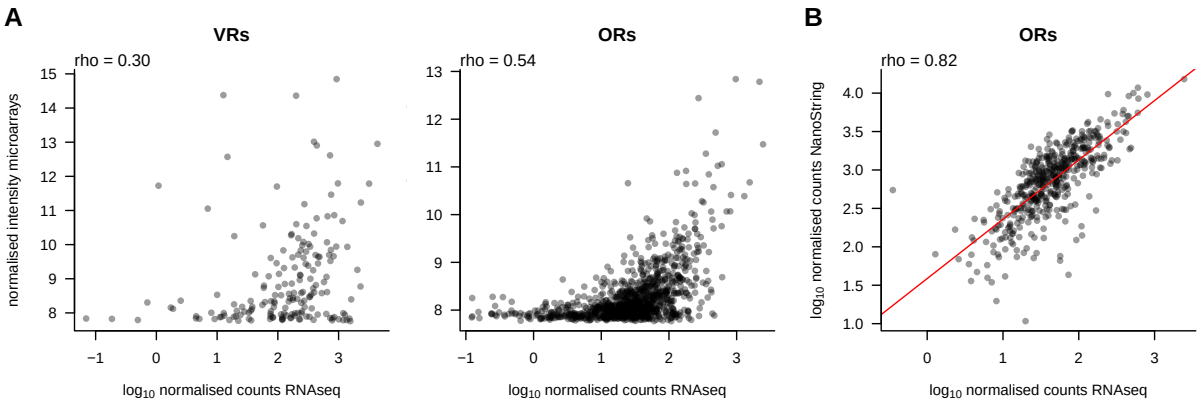


Figure 2.9 – Comparison of the receptor expression across different platforms. A) Scatter plots of the expression values for VR and OR genes in the RNAseq data versus the microarrays. Many receptors have intensity values around 8 in the arrays, the threshold between signal and noise; these genes are scattered along several orders of magnitude in the RNAseq data. **B)** Comparison of the expression values for OR genes in the RNAseq data versus the NanoString nCounter estimations. Values are highly correlated. In red is the regression line. For all, the Spearman's correlation coefficient is indicated in the top left corner.

Khan et al. profiled 531 OR genes in C57BL/6 animals, using NanoString nCounter [304]. When compared to the RNAseq expression levels, the NanoString normalised counts are highly correlated (rho = 0.82, p-value < 2.2e-16; Figure 2.9B). The agreement of these two technologies, which are based on very different detection principles, provides further support for the accuracy in the quantification of expression of receptor genes by RNAseq.

2.2.2 Sensitivity of RNAseq to detect lowly expressed receptor genes.

The expression of the majority of the receptor genes is very low, but they still display a dynamic range of expression levels (inset in Figure 2.7) that are consistent between biological replicates (0.81 < rho < 0.92, p-value < 2.2e-16). At such low expression values, the variance of the data increases considerably, but the high correlation coefficients imply that the relationships between the different receptors are preserved. To further test whether this range of expression values is meaningful, I analysed RNAseq data obtained from the WOM of mice with a homozygous deletion of the second cluster of OR genes in chromosome 9 (S. Xie, P. Feinstein, and P. Mombaerts, unpublished data; [306])[3]. Of the 94 deleted OR genes within the cluster, 83 (88.3%) had no counts in any of three biological replicates; importantly, these same genes show expression estimates up to 301 normalised counts in control animals (Figure 2.10). The 11 remaining genes have just one or two fragments mapped in only one of the replicates, resulting in normalised counts of less than 0.4.

This suggests that RNAseq suffers from virtually no noise and that very low expression levels reflect real expression in the samples profiled. Therefore, I propose that the RNAseq methodology presented here is able to detect the expression of most of the putative functional receptors in complex samples of whole tissue from the olfactory system, despite their low expression levels. This provides, for the first time, almost complete evidence of expression for the computationally predicted receptor genes in their cognate tissue, which supports their role in olfactory signalling.

2.2.3 The multiread problem.

The major limitation of RNAseq is that it is based on short fragments; these are not always unique in the genome, which means that by reading their sequence, one cannot

[3]RNA from these animals was kindly obtained and provided by Mona Khan.

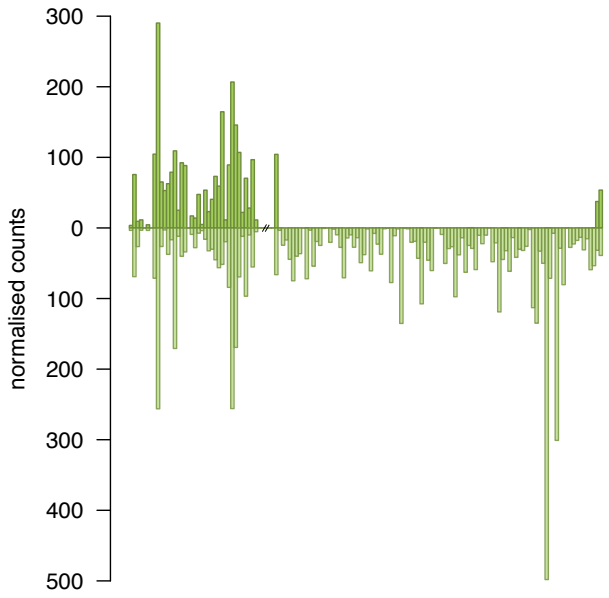


Figure 2.10 – Expression of a deleted OR cluster. The mean normalised expression for each receptor in chromosome 9 is plotted for a mouse with a homozygous deletion of an OR cluster (dark green); the break in the x-axis separates the two clusters present in this chromosome. While expression is detected for the first cluster, the OR genes from the deleted cluster are not expressed, except for the ones at the borders, which were not included in the deletion. The corresponding expression in wild-type animals is plotted as a mirror image (light green).

know their true place of origin. These multireads occur much more often in genes with close paralogs and may, therefore, represent a big problem for accurately estimating the expression of receptor genes. To assess this, I first calculated the *uniqueness* of each receptor gene with different lengths of reads; that is, I counted how many of all the possible n-mers obtained from a given receptor gene can be mapped unambiguously to the genome, with $n = (32, 76, 100)$ nucleotides (nts), which are common read lengths.

In general, OR genes are much more unique than VR genes (p-value $<$ 2.2e-16, Mann Whitney test). Sequence fragments of 32 nts are very often multimapped; at this fragment length, 43.7% and 15.8% of the VR and OR genes respectively have less than 50% uniqueness. However, a steep increase in the uniqueness values occurs when the sequence length is increased to 76 nts and then a modest gain when fragment length reaches a 100 nts (Figure 2.11). These values, however, are a worst-case scenario, since paired-end reads will improve the uniqueness of many genes. Therefore, RNAseq based on reads of 76 or 100 nts should be able to differentiate between most receptors. Nonetheless, even at 100 nts, 53 (9.4%) VR genes have uniqueness of 0, and 105 (18.6%) have uniqueness $<$ 25%; these are virtually inaccessible for any current profiling technique.

Despite the ability to detect expression of most receptor genes, those with low unique-

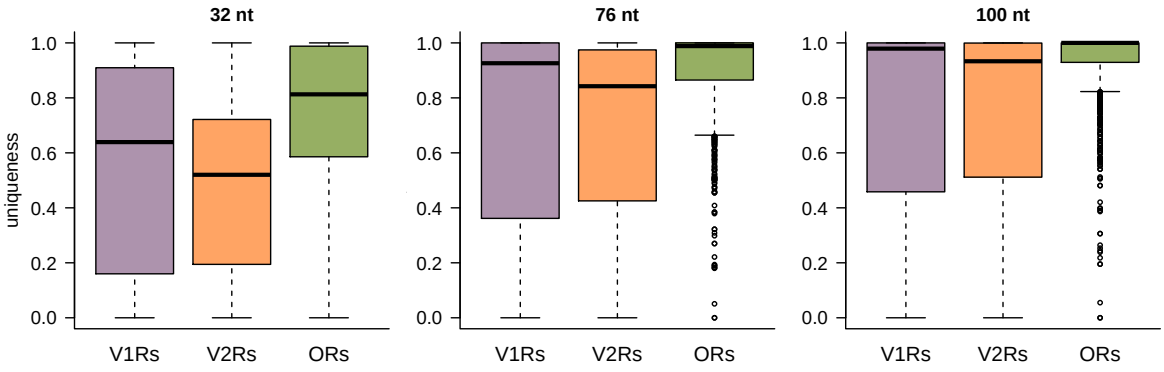


Figure 2.11 – Uniqueness of the receptor sequences. Boxplots indicating the uniqueness values for V1R, V2R and OR genes, for read length of 32, 76 and 100 nucleotides long.

ness will be consistently underestimated. I therefore analysed how much information is lost in multireads and whether this affects some classes of receptors more than others. For each receptor, I counted the number of different sequencing fragments that map to it; therefore, each multiread was counted at least twice, once for each receptor it maps to. On average, 37% and 94.1% of the multi-mapped fragments had only two alignments, for VR and OR genes respectively, with the rest mapping to three or more different locations. These counts were then normalised to account for depth of sequencing and gene length, and compared to the corresponding unique counts (Figure 2.12).

Consistent with the uniqueness calculations, there are many more multireads assigned to VR than to OR genes, and there are some receptor families that have many more multireads than others. For example, there is a cluster of 84 V1Rs in chromosome 7 that have very low unique counts (median = 0.1), and a low but consistent number of multireads (median = 50.01); accordingly, the uniqueness of these genes is very low (median = 0; mean = 0.12). For the VR repertoire, 44.85% of the receptors have more multireads than unique counts; in contrast, for the OR genes, this is the case for only 9.9%. Therefore, the expression estimates of some classes of VR genes are not accurately accounted by RNAseq, but the majority of the OR repertoire is not significantly affected by this problem.

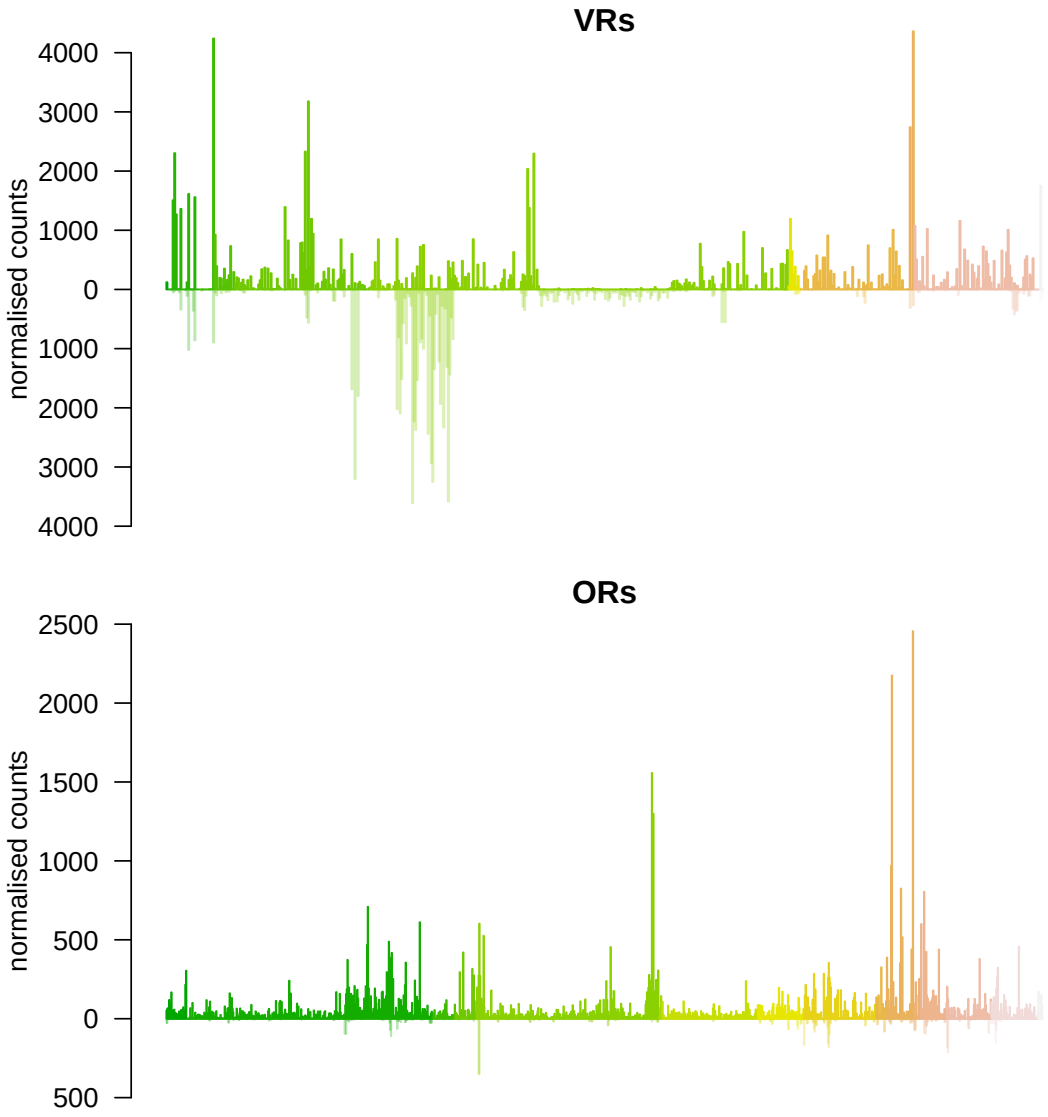


Figure 2.12 – **Multireads mapped to VR and OR genes.** Barplot of the normalised counts for each VR (top) and OR (bottom) gene; receptors are ordered according to their chromosomal location and each chromosome is a different colour. As mirror images are the corresponding normalised counts of multireads for each receptor.

2.2.4 Complete annotation of the gene models.

An advantage of RNAseq over other expression profiling techniques is that it is not restricted by a catalog of known transcripts. Using the sequencing data directly, it is possible to reconstruct the most parsimonious set of transcripts that would generate the observed reads. This can be used to annotate new transcript isoforms for known genes, or to identify completely novel genes. For most of the receptor genes, only the CDS is annotated, based on computational predictions. However, there is evidence that at least some receptors contain additional non-coding exons and complex UTRs that undergo alternative splicing[62]. I used Cufflinks[307] to generate *de novo* assemblies in order to identify full length transcripts for OR and VR genes. Samples were sequenced at sufficient depth to produce new, extended receptor gene models for 913 (73%) OR and 246 (43.7%) VR genes. I identified additional exons for many of the receptor genes: 866 and 68 OR genes have exons 5' and 3' to the CDS, respectively; and 163 and 79 VR genes have exons 5' and 3' to the ORF (Figure 2.13).

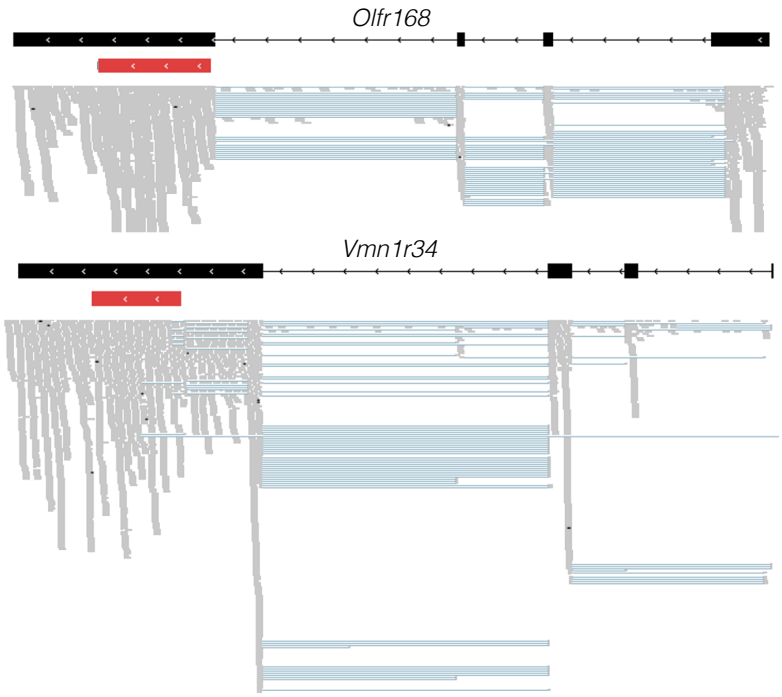


Figure 2.13 – Full length gene models for OR and VR genes. Examples of the generated gene models for an OR (top) and a V1R (bottom) are shown in black. Boxes represent exons and the lines joining them are introns; the arrow heads indicate the strand of the gene. The existing Ensembl annotations for the genes are shown in red with their UTRs in grey. Underneath, the sequencing reads are shown in grey; those that span exon-exon junctions are joined by blue lines.

OR and V1R genes typically have coding regions that span a single exon, but I identified 54 OR and 15 V1R genes where at least one of the reconstructed transcripts

has an intron within the protein coding sequence (as annotated in Ensembl). The predicted ORFs for most of these transcripts are truncated due to a premature stop codon. But for 17 OR and 3 V1R genes the ORF is of typical length, and could encode a putatively functional receptor. All but one (*Olfr332*) are annotated in Ensembl and classified as protein coding.

To investigate cases of alternative splicing, I retained all the multi-exonic receptor gene models and counted the number of alternative isoforms produced. 70% of VR genes have between 1 and 4 isoforms while 85% of OR genes have 1 to 3 isoforms (Figure 2.14). A few receptor genes have more than 8 different isoforms (38 VR and 10 OR genes) but in most of these cases isoforms have alternative transcription start sites (TSS) or exons that differ in length by just a few nucleotides; therefore, several of the final transcripts differ only very slightly.

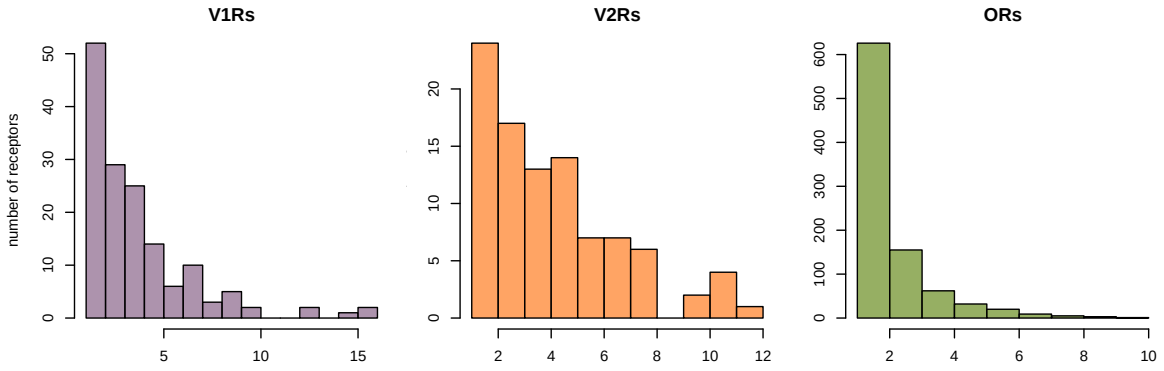


Figure 2.14 – Number of transcripts per receptor gene. Histograms of the number of different transcripts isoforms for each V1R, V2R and OR gene.

Next, I calculated the length for each receptor gene based on the existing Ensembl and the new reconstructed models. The median length for both the Ensembl OR and V1R gene models is about 950 nts, while the corresponding reconstructed gene models are now around 2,500 nts long. The median length of Ensembl V2R genes is 2,559 nts, while for the corresponding reconstructed gene models is 2,912 nts (Figure 2.15A). The increase in length is dominated by a long 3' UTR. UTR sequences are more variable than the CDS[299, 300] and, therefore, could help to differentiate between highly similar OR and particularly VR transcripts. I therefore assessed the uniqueness of these new gene models and found a large increase in the proportion of unique sequence for the V1R (p-value < 0.0001, Mann Whitney test) and V2R gene models (p-value < 0.0001, Mann Whitney test); a more modest increase was apparent in OR genes (p-value $= 0.044$, Mann Whitney test; Figure 2.15B).

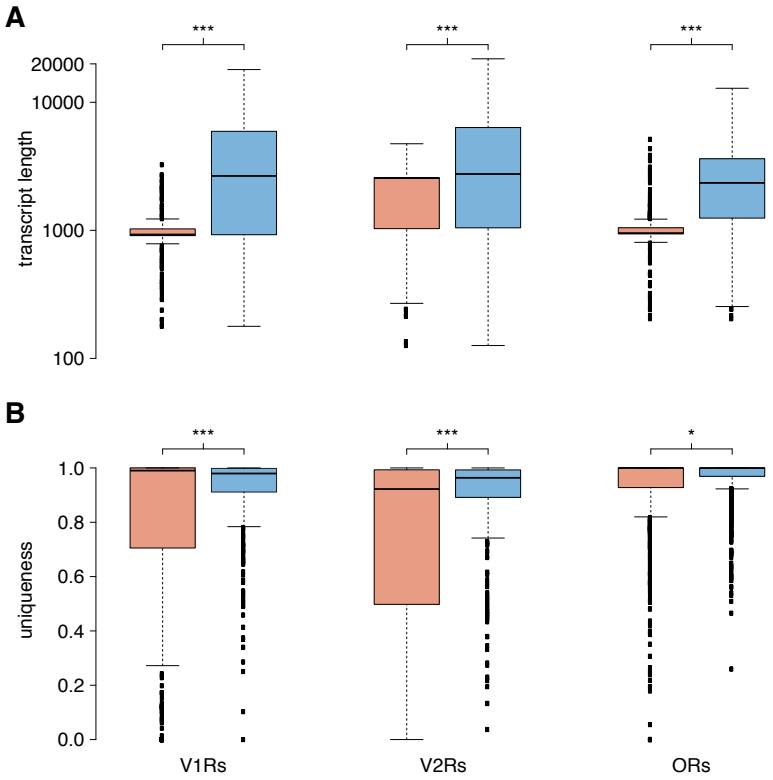


Figure 2.15 – Improved receptor gene models are more unique. **A)** Boxplots of the transcript length in Ensembl (orange) or the reconstructed gene models (blue) for the V1R, V2R and OR genes. The increase in transcript length is highly significant (*** P<0.0001, two-tailed Mann-Whitney test). **B)** As above, but quantifying the proportion of unique sequence (*** P<0.0001 and *P<0.01, two-tailed Mann-Whitney test). The uniqueness corresponds to the proportion of all 100 nucleotide long windows within the transcript that map uniquely to the genome.

Finally, I compared the 5' ends of the OR gene models reconstructed from the RNAseq data to the proposed TSS reported by Plessy et al. using nanoCAGE[193]. A third of the OR genes differ in less than 20 nucleotides, and almost 85% are within a 500 nucleotide window. However 34 OR genes have a discrepancy of more than 5 kb, where the 5' end proposed by nanoCAGE is upstream of the one found by Cufflinks. A close examination of the sequencing data for the 25 genes with the biggest 5' differences revealed that, for 24 genes, there were no sequencing fragments consistent with the TSS proposed by nanoCAGE. In 12 of these cases, the nanoCAGE TSS overlaps with the 3' UTR of an adjacent OR gene and 2 represent the TSS of a different gene. Only one TSS is correctly inferred by nanoCAGE where Cufflinks failed to reconstruct the full-length model. Clowney et al. also defined the 5' end of OR genes using tiling microarrays[194]. A similar comparison analysis revealed that a third of the receptor genes differ in less than 100 nucleotides and 80% of the data is contained within a 1.5 kb

window. Therefore, the reconstructed gene models based on the RNAseq data largely agree with previous reports, but provide a much better resolution and a detailed exon structure that has never been reported for such a large proportion of the receptor repertoire. The defined UTR sequences should aid in the design of more specific probes for experiments based on hybridisation strategies.

To assess how much additional information is provided by the improved annotation of the receptor repertoire and their increased uniqueness, I re-estimated the expression using the reconstructed gene models. The count data was normalised to account for depth of sequencing but not for gene length, since this is different for the new models and some receptors increased more than others. Overall, for those genes with an extended gene model, there is a clear increase in the number of sequencing fragments recovered. On average, expression estimates are 4.4, 1.4 and 1.9 fold more abundant for V1R, V2R and OR genes respectively (Figure 2.16). These additional data should improve the sensitivity and accuracy of the estimated expression levels, and some genes that were previously inaccessible now have unique sequencing fragments mapped to their UTR regions.

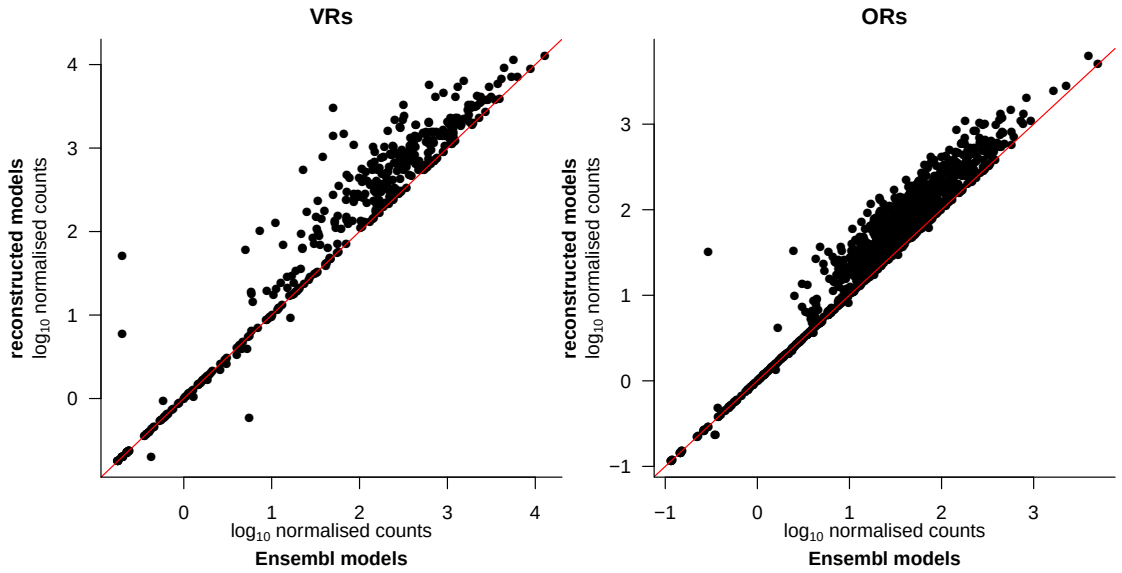


Figure 2.16 – Expression of the receptors with the new gene models. Scatter plot of the normalised expression estimates for each VR (left) and OR (right) gene using the Ensembl gene models (x-axis) versus the reconstructed gene models (y-axis). The red line indicates the 1:1 diagonal.

2.3 Identification of novel genes.

I then extended the analysis of the *de novo* assembly performed by Cufflinks to the full olfactory transcriptomes. This revealed 5,562 and 6,228 loci that have evidence of transcription in the VNO and WOM respectively, that do not overlap any annotated genes in the Ensembl database. 40% of these loci are found in both tissues. Many are located in close proximity to the start or end of annotated genes and are likely to represent unannotated UTRs. To search for new genes, I first excluded all those predictions that lie within 5 kb of cataloged genes. Of the remaining features, about 75% represent single-exon transcripts, leaving 756 and 847 putatively novel multi-exonic genes expressed in the VNO and WOM respectively (Table 2.1).

	VNO	**WOM**
Total identified genes	**5,562**	**6,228**
Shared between tissues	2,331	2,519
Within 5kb of annotated genes	2,564	2,889
Putative novel gene models	**2,998**	**3,339**
Single-exon predictions	2,242	2,492
Multi-exon predictions	**756**	**847**
Aligned to a protein domain	229	258
Overlap with a predicted transcript	625	694

Table 2.1 – Putative novel genes. Number of novel genes predicted from the RNAseq data (that do not overlap any annotated gene in Ensembl). Predictions within 5 kb of annotated genes were excluded and the remaining are considered putative novel gene models. Putative genes are considered shared between tissues if 50% or more of the gene length is found in both the VNO and WOM. In some cases two predicted genes in one tissue can overlap with a single prediction in the other, leading to a different number shared in each.

Ensembl provides databases of annotated genomic regions that match protein domains or computationally predicted genes. When cross-referenced to the putative novel genes, about 30% had matches to known protein domains and 80% overlapped with *in silico* predicted transcripts, suggesting that these might be true genes. Some of them are surprisingly abundantly expressed. One of the genes reconstructed from the VNO data is the 6th most abundant gene in the transcriptome, and it is also present in the WOM samples though at much lower levels. Interestingly, there is a second predicted gene in close proximity with a similar exon-intron structure (Figure 2.17A). These two genes were validated in collaboration with members of the lab. Full-length transcripts were cloned for both of these genes and we could identify ORFs on opposite strands that encode two closely related proteins; by homology, these could be classified as novel

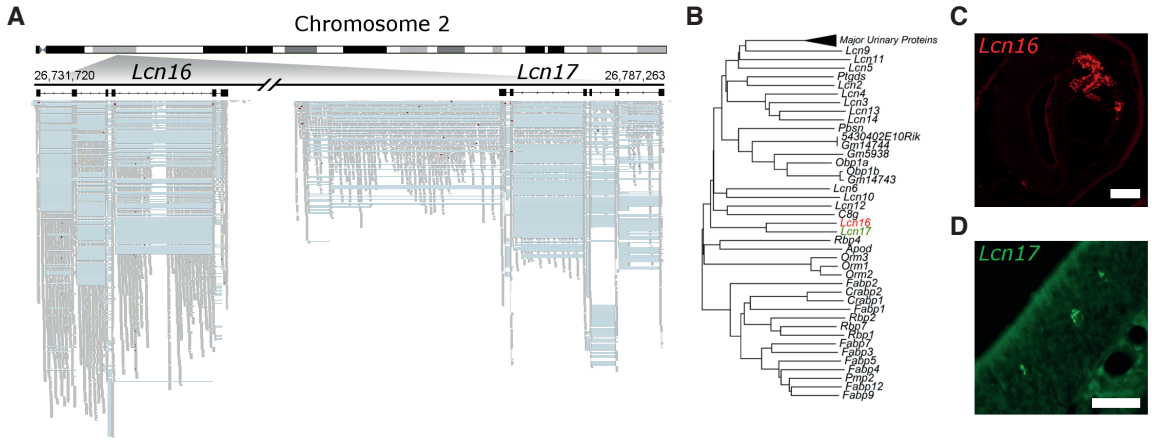


Figure 2.17 – Novel genes expressed in the olfactory system. A) Chromosome 2 is schematized on the top and the locus where two previously unidentified genes were found is amplified below. In black are *Lcn16* and *Lcn17* gene models; boxes correspond to the exons and lines joining them represent introns. The mapped RNAseq reads are below: each read is drawn in grey and blue lines join read fragments that span exon-exon junctions. **B)** Phylogenetic reconstruction of the novel genes with other members of the mouse lipocalin gene family. **C)** *In situ* hybridisation reveals *Lcn16* is expressed in glandular tissue of the VNO and **D)** *Lcn17* is expressed in cells within the MOE. Scale bars: (B) 100 mm, (C) 50 mm.

members of the lipocalin gene family, and were named *Lcn16* and *Lcn17*[4]. A phylogeny of all mouse lipocalins revealed these genes form a distinct sub-clade (Figure 2.17B). *In situ* hybridisation experiments confirmed that *Lcn16* is expressed abundantly in glandular tissues of the VNO, while *Lcn17* is expressed in a small number of cells in the MOE (Figure 2.17C-D)[5]. Despite the mouse genome being one of the best annotated genomes available, this data suggests that there still exists a significant number of unannotated genes. Since the olfactory system is not a tissue routinely used for gene identification, many of these novel genes are likely to have olfactory-related functions and, therefore, would have been missed in other tissues. The validation of two putative genes suggests that many more might be true uncharacterised genes.

Overall, I have shown that RNAseq not only is the best technology available to study the transcriptome of the olfactory system, but it provides reproducible, accurate information about its transcriptome. I have produced a comprehensive profile of the genes expressed in both the VNO and the WOM, that represent the average expression across all the different cell types present therein. Furthermore, I have identified additional genes that have never been characterised before. Most importantly, these data has demonstrated great specificity in the detection of the receptor repertoires which,

[4]The sequences for *Lcn16* and *Lcn17* are available in GenBank under accessions KJ004569 and KJ004570.

[5]The full-length transcripts were obtained by Maria Levitin; the phylogeny was produced by Darren Logan; the *in situ* hybridisation was performed by Luis Saraiva.

until now, have been very difficult to study as a whole. I have demonstrated that the reconstruction of full-length gene models for a significant proportion of the receptor genes allows better estimation of their expression values and increases significantly the amount of unique sequence available for each of them. Some subfamilies of VR genes still represent a challenge, because they are composed of nearly identical genes. These will be very difficult to accurately differentiate with nearly every technique available to date. Nonetheless, a good proportion of the VR repertoire, and nearly the complete OR repertoire are divergent enough across their full length, for accurate discrimination and further study.

Chapter 3

Decomposing the WOM: from tissue to single-cell.

The material presented in this chapter has been previously published in reference[308]. I confirm I have ownership of the copyright for its reproduction. All the work presented here has been done in collaboration; the experiments that led to the collection of cells and tissues were performed by others, but I have solely analysed all the data generated. The results presented are my own work unless otherwise stated.

The data presented in the previous chapter provides a generalised profile of the transcriptome of the two most prominent and better studied components of the mouse olfactory system. Such gene catalogues, however, represent the averaged expression of each gene across the different cell types found within each tissue. The sensory neurones are the primary focus in this dissertation, given their role in olfactory signalling, and particularly the receptor repertoire. Therefore, I established a collaboration project with Peter Mombaerts and Mona Khan to specifically isolate the OSNs of the MOE. For this, a transgenic mouse was used, that is engineered to express GFP from the OMP locus (OMP-GFP)[309]. OMP is a protein specifically expressed in the sensory neurones from the MOE and VNO[310], and it is not expressed in sustentacular or basal cells[311]. Its expression can be identified after both OR genes and *Adcy3* have been activated. Additionally, the expression domain of *Omp* doesn't overlap with *Gap43*, a marker of immature OSNs. Therefore, OSNs that express OMP are considered mature[312]. Analysis of an OMP KO animal revealed that the general structure of the MOE is unaffected and

the ratio of mature to immature neurones also remains normal; but the MOB is 15% smaller. EOG recordings upon stimulation with several odours resulted in decreased recordings, with slower kinetics for both the initiation and recovery of the response. Nonetheless, OSNs were still able to be activated and mice were not anosmic[313]. The OMP-GFP mouse is a knock-in, where the CDS of *Omp* was substituted for GFP. To avoid altering the response kinetics of the OSNs, all experiments were conducted in heterozygous animals.

3.1 The transcriptome of the olfactory sensory neurones.

To characterise the transcriptome of the OSNs alone, dissociated cells from the WOM of heterozygous OMP-GFP mice were separated based on their expression of GFP, by fluorescence-activated cell sorting (FACS). The GFP^+ –and therefore OMP^+– cells were retained, and the populations from several animals were pooled to obtain 10 million cells. This is the number of mature OSNs typically found in a single adult mouse[224]. Three biological replicates were collected and the RNA extracted from the OSNs was used to construct libraries for RNAseq (Table B.1 in Appendix B). Since the OMP-GFP mouse is in a mixed genetic background (129P2 X B6), three WOM samples from single animals were also processed for comparison[1]. All six samples were sequenced at high depth on the Illumina platform and data was analysed including the extended models for receptor genes (Table B.2 in Appendix B).

As observed previously, biological replicates were highly correlated (rho $>$ 0.96, p-value $<$ 2.2e-16) for both the WOM samples and the pooled OSNs. In order to reveal genes that are preferentially expressed in the OSNs or in other cell types, I performed a differential expression (DE) analysis. 67.6% of all the genes expressed were significantly different between the OSNs and the whole tissue (FDR $<$ 5%), with 45.8% of these being expressed more abundantly in the OSNs (Figure 3.1). From these, 790 have a fold-change greater than 3 and 50.1% are OR and TAAR genes. To explore their functions, I performed a gene ontology (GO) analysis that revealed enrichment for terms related to the olfactory transduction pathway and G-protein coupled amine receptor activity, along with genes related to synaptic vesicles, branching morphogenesis of a nerve and peptide hormone processing. In contrast, 5,227 genes were expressed higher in the WOM

[1]The collection and RNA extraction from WOM samples and pools of FAC-sorted OSNs were performed by Mona Khan.

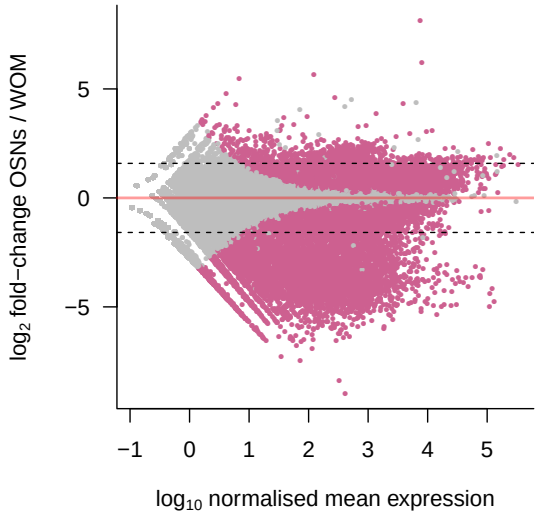


Figure 3.1 – Differentially expressed genes between the OSNs and WOM. MA plot showing the mean expression for each gene against its corresponding fold-change value between the OSN and WOM samples. The red line represents equal expression in both groups. Significantly DE genes are in pink (FDR < 5%). Dotted lines represent a threshold of fold-change of ± 3.

(fold-change < 0.33, FDR 5%), and over half of these were expressed at least ten times higher than in the pooled OSNs; this suggests that they are likely restricted to the non-neuronal cell types found within the WOM samples.

A similar study was performed by Sammeta et al.[314] using microarrays, to compare the transcriptomes of the GFP^+ versus the GFP^- populations of an OMP-GFP mouse. The genes that were reported as enriched in the OMP^+ population in this dataset tend to be more abundant in the pooled OSN samples; and the genes expressed specifically in the WOM are classified as enriched in the OMP^- population (Figure 3.2), thus supporting the accuracy of the DE analysis results.

Next, I compared the expression of the receptor repertoire. As expected, both OR and TAAR genes were generally expressed at higher levels in the OSNs than in the WOM. There was a median fold-change increase of 2.56 in overall expression, and a similar increment is observed for the canonical markers of mature OSNs (Figure 3.3). The overall expression values were highly correlated (rho=0.95, p-value < 2.2e-16), which indicates that the repertoire's expression increases as a whole, maintaining the proportions between receptors observed in the whole tissue. Only 19 OR genes that are present in the WOM samples are lacking in the OSN pools. However, all these are expressed at very low levels and 13 (68.4%) are annotated pseudogenes. As observed previously, nearly the complete repertoire of OR genes is expressed, as assessed by unique counts. Over 95% of the OR genes have at least one fragment mapped in both the WOM and

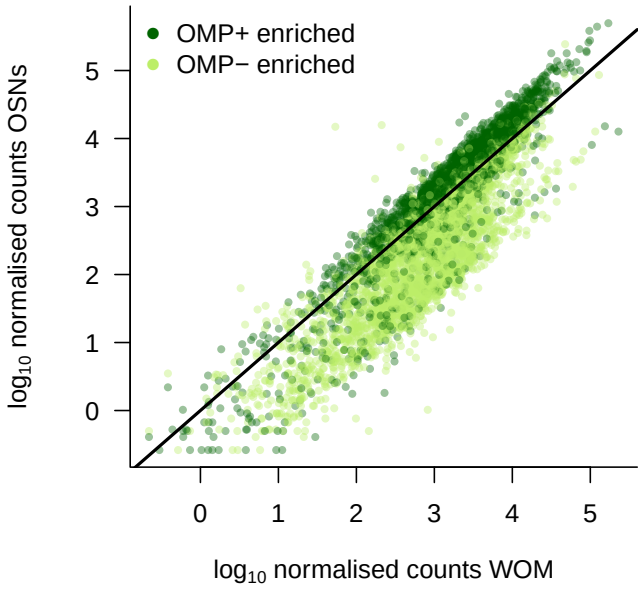


Figure 3.2 – **Comparison to Sammeta et al.** Scatter plot of the expression values in the WOM versus OSN samples. Genes are coloured depending on whether Sametta et al. classified them as enriched in the OMP^+ cells (dark green) or the OMP^- cells (light green). The line indicates the 1:1 diagonal. Genes above the diagonal are expressed higher in OSNs and tend to be enriched in the OMP^+ sample, whereas genes below the diagonal are expressed higher in the WOM and are enriched in the OMP^- sample.

OSN samples; this number increases to 98.9% if only the ORs annotated as functional are considered.

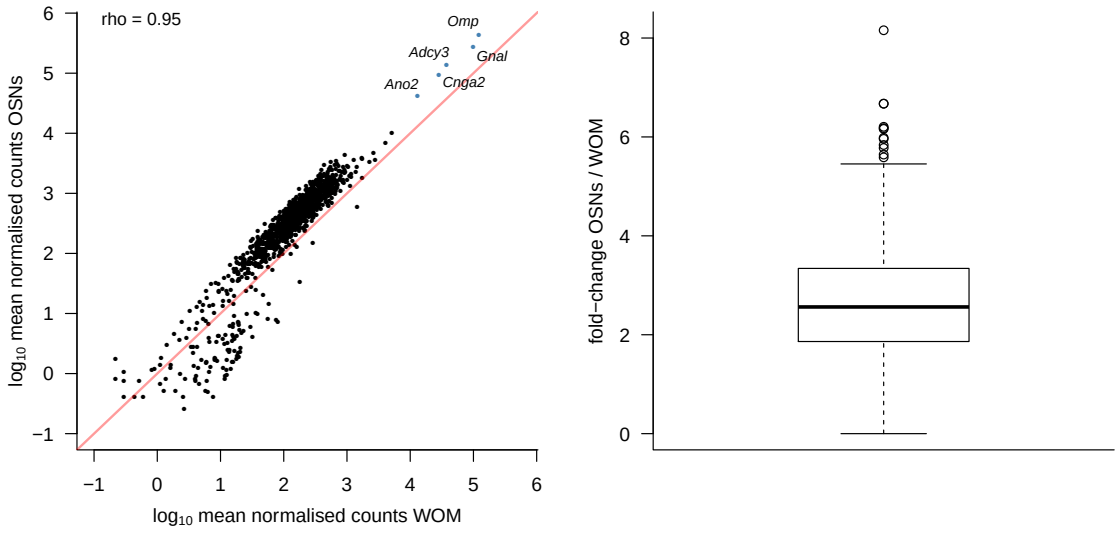


Figure 3.3 – **Receptor expression in the WOM vs the sorted OSNs.** On the left, scatter plot of the expression of all OR and TAAR genes (black) in the WOM versus the sorted OSNs, and of five canonical markers of mature OSNs (blue). The red line represents the 1:1 diagonal. The majority of genes are expressed more abundantly in the sorted OSNs. On the right, a boxplot of the fold-change between the OSNs and the WOM for the receptor expression levels. The median increase in expression is 2.56.

All together, I have characterised the transcriptome of the neuronal component of the MOE, revealing many genes that are preferentially expressed in OSNs. Additionally, thousands of genes can now be classified as specific to the other cell types of the MOE and surrounding tissue. The deep RNAseq strategy used is effective in characterising the expression of the whole receptor repertoire in this transgenic mouse strain also. By separating the OSNs from all the other cell types, the depth of sequencing devoted to receptor genes is increased, but the proportionality of the different neuronal subpopulations present within the MOE is conserved. The very high correlation between the expression estimates for the receptors in both sample types argues in favour of their accuracy and reproducibility.

3.2 Mature OSNs segregate into two distinct populations.

When cells were dissociated from the WOM of OMP-GFP animals and sorted based on their GFP intensity, it was evident that two different populations of GFP^+ cells are present in the samples (Figure 3.4). Cells from one population have much more intense fluorescence levels than the others, so I will refer to these as the GFP^{high} and GFP^{low} populations. To investigate further this distinction, pools of 10,000 cells from each population were obtained from three different animals for RNAseq (Table B.1 in Appendix B)[2]. The expression estimates between replicates were highly correlated ($0.87 < \text{rho} < 0.91$, p-value $<$ 2.2e-16) despite the smaller number of cells used for RNA extraction. However, one of the GFP^{high} samples yielded only a small number of sequencing fragments compared to the others, and it was excluded from downstream analyses (Table B.2 in Appendix B).

As a first control, I examined the levels of *Omp* expression in each population. In all samples there was clear, robust expression of this marker gene, but the GFP^{high} samples had a consistent 1.52 fold increase in expression (p-value = 0.02482, t-test; Figure 3.5A). All the GFP^+ cells should have the transcriptional profile of mature OSNs, since they robustly express *Omp*. To confirm this, I compared the RNAseq data profiles to a microarray dataset that characterised 670 and 565 genes as preferentially expressed in mature and immature OSNs, respectively[315]. Both the genes enriched in mature and immature OSNs were expressed at similar levels in both the GFP^{high} and GFP^{low}

[2]Sorting experiments, RNA extraction and library preparation were performed by Luis Saraiva.

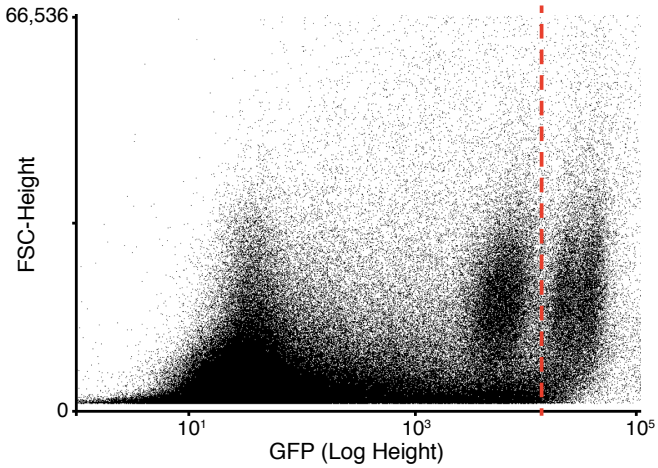


Figure 3.4 – FACS plot of OMP-GFP animals. Each dot is a cell plotted according to its GFP fluorescence intensity (x-axis). For the GFP$^+$ cells, two populations are visible. The red dotted line roughly separates them. Image kindly provided by Luis Saraiva.

populations (p-value = 0.59, t-test), but the genes from mature OSNs were expressed significantly higher than the ones from immature neurones (p-value = 1.73e-07 for GFP^{low}, and p-value = 1.35e-07 for GFP^{high}, t-test; Figure 3.5B).

To identify other genes that differentiate these two populations of OSNs I performed a DE analysis. This revealed 537 significantly DE genes (FDR 5%), 420 (78.2%) of

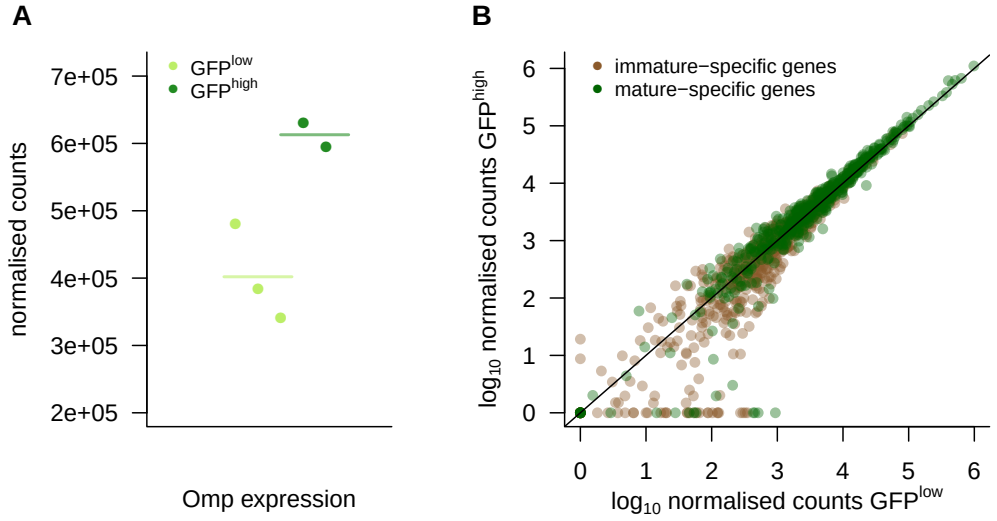


Figure 3.5 – GFP$^+$ neurones are mature. A) Both the GFP^{low} and GFP^{high} populations express high levels of *Omp*, but the GFP^{high} samples have 1.52 times higher expression. **B)** Scatter plot of the normalised expression in the GFP^{low} versus the GFP^{high} samples. The line indicates the 1:1 diagonal. Genes are coloured according to whether they were characterised as enriched in mature (green) or immature (brown) OSNs. All genes are expressed at similar levels in both populations, but the genes from mature OSNs are expressed at higher levels than those from immature neurones.

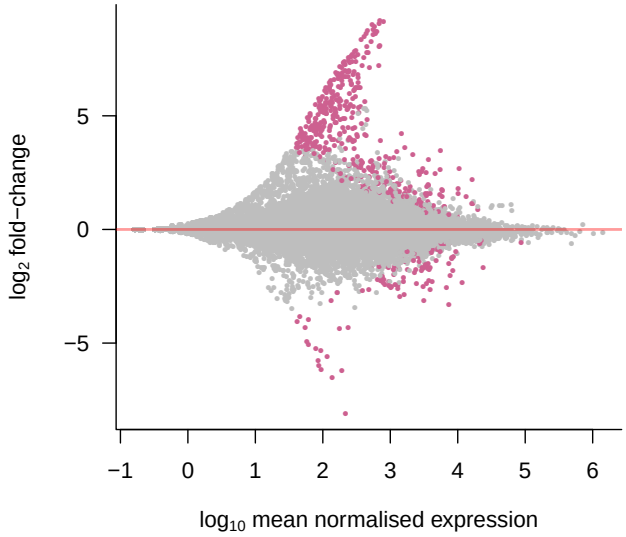


Figure 3.6 – Differentially expressed genes between the GFP^{low} and the GFP^{high} cells. MA plot showing the mean expression for each gene against its corresponding fold-change value between the GFP^{low} and the GFP^{high} samples. The red line represents equal expression in both samples. Significantly DE genes are in pink (FDR $< 5\%$).

which were more highly expressed in the GFP^{low} population (Figure 3.6). This set of genes is enriched for terms related to development, morphogenesis, negative regulation of neuronal differentiation and positive regulation of cell proliferation. Therefore, it appears that the different levels of *Omp* correlate with the level of maturation of the cells. The GFP^{low} cells are still in the process of downregulating genes involved in proliferation and have yet to achieve final differentiation.

These data indicate that there exists a previously unknown subdivision of mature OSNs characterised by the levels of *Omp* expression. While both populations have a characteristic expression profile of mature OSNs, the neurones from the GFP^{low} population are less mature than the rest.

3.3 RNAseq of single OSNs.

I have demonstrated that the mouse MOE is composed of over a thousand different subpopulations of OSNs, each characterised by the OR gene they express. We were interested to know how homogeneous are the transcriptomes of neurones from these different subpopulations. To this end, a Fluidigm C1 microfluidic system was utilised to

capture single OSNs from the FAC-sorted GFP^{high} population from one mouse[3]. This system uses a fluidic circuit to isolate single cells into individual reaction chambers, that can be examined under the microscope to ensure that a single cell is present. After capture, the system performs all the necessary reactions to produce cDNA from each cell (cell lysis, mRNA reverse transcription and PCR amplification) that can then be used to prepare libraries for RNAseq (Table B.4 in Appendix B). From the 96 wells in the capture chip, 58 contained single cells, 8 were empty and the remaining contained more than one cell and/or had visible debris contamination (Figure 3.7A). Only single cells were considered in further analyses.

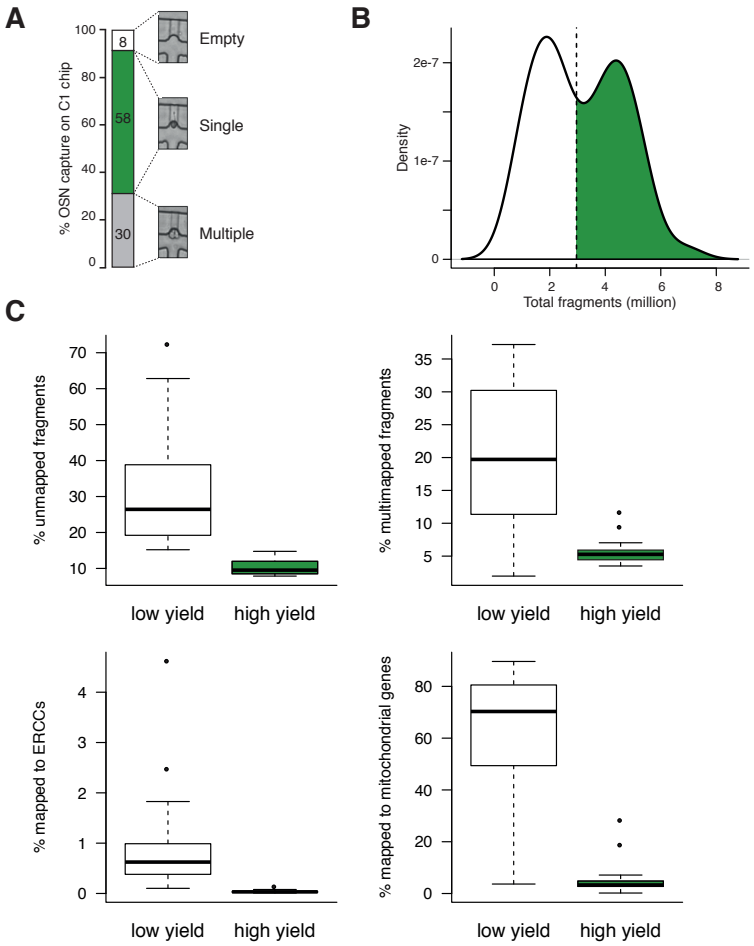


Figure 3.7 – **Quality control of the single-cell RNAseq data. A)** Proportion of wells in the C1 capture chip that were empty (8), contained one cell (58) or more than one cell and/or debris contamination (30). Representative bright field images of each case are shown. **B)** Density distribution of the total number of sequencing fragments obtained for the 58 single-cell samples. Decomposition of the distribution into two normal-like distributions splits the data at 2.96 million (dotted line), separating the samples into those with low (white) and high (green) yield. **C)** Several mapping statistics are clearly different for the low and high yield samples. Schematic in A was kindly provided by Luis Saraiva.

[3]All the capture and library preparation experiments were performed by Luis Saraiva.

The quality of the RNAseq data produced from a single cell is greatly influenced by the amount and integrity of the starting material. Including cells with poor quality data in normalisation steps can be detrimental to the downstream analyses of cells with good quality data; therefore, it is imperative that these are identified and excluded at an early stage[316]. The distribution of total fragments obtained from each of the 58 single-cell samples was clearly bimodal (Figure 3.7B). Deconvolution into two normal-like distributions revealed 28 cells with low yield (mean of 1.7 million) while the remaining 30 were sequenced at significant higher levels (mean of 4.4 million; p-value $<$ 2.2e-16, t-test). To determine if the lower yield was a result of sequencing poor quality libraries, I analysed the mapping statistics (Figure 3.7C). The low-yield group of samples had a much higher percentage of unmapped fragments (31.97% on average versus 10.21% in the high-yield group; p-value = 5.739e-08, t-test) as well as multi-mapped fragments (20.27% versus 5.49%; p-value = 4.018e-08, t-test). The proportion of fragments that mapped to ERCC spike-ins was over 20 times higher in the low- versus high-yield groups (p-value = 5.576e-05, t-test). Furthermore, most of the uniquely mapped fragments from the low yield samples aligned to mitochondrial genes (on average 60.53%, compared to 4.77% in the high-yield samples; p-value = 6.327e-12, t-test). Together these suggest that the starting material for the samples with low yield was of poor quality and were therefore excluded. I focused subsequent analyses on the 30 high-yield samples.

Next, I compared the OR gene expression profiles of these samples as a function of their capture location on the C1 capture chip and the library preparation plate. Ten cells (OSN 157, 178, 185, 191, 207, 214, 218, 223, 255 and 263) had evidence of two highly expressed OR genes. OSN 263 had high counts for *Olfr55* and *Olfr239*, two adjacent OR genes that are 99% identical. Closer inspection of the sequencing data revealed that the fragments assigned to *Olfr239* were in fact mismapped, since a BLAST alignment mapped them back to *Olfr55*. Therefore, I set the counts of *Olfr239* to zero. For the remaining nine cells, in four cases (OSN 178, 191, 223 and 255) I found that one of the OR genes was expressed in another sample located in the immediately adjacent well, suggesting evidence of carry-over. The other five cells (OSN 185, 207, 214, 218 and 257) did not share an OR gene with a sample in an adjacent well. I independently reassessed these nine cells through all previous quality control criteria and could not distinguish the four cells with evidence of carry-over from the five cells without it. A recent report found that up to 20% of cells captured on a C1 microfluidic system contain two cells that are not visible in the microscopy images[317]. Thus, to take a conservative approach and to reduce the possibility of including samples containing a second, visually obscured cell

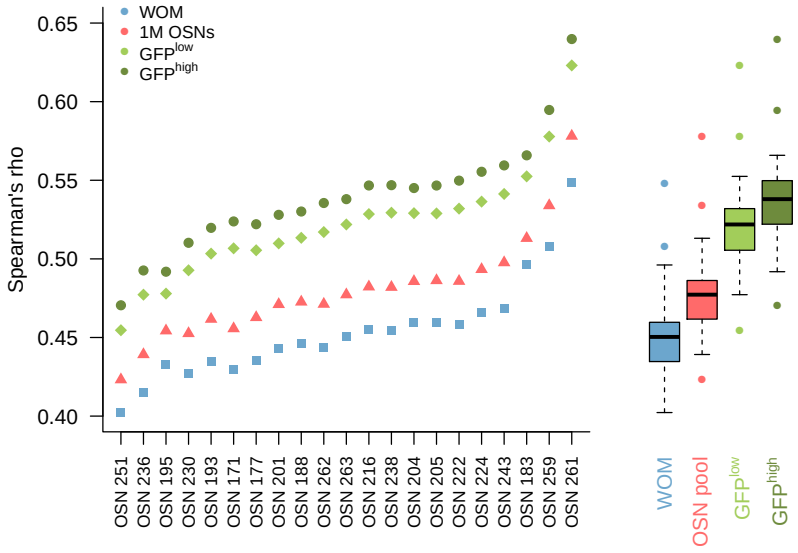


Figure 3.8 – Correlation of single OSNs to other datasets. The Spearman correlation coefficient is plotted for each single OSN (x-axis) against the previous population datasets. The single cells correlate better with the isolated OSNs that with the WOM, and better with the GFP^{high} than the GFP^{low} neurones. On the right, the cumulative data for each comparison is presented as boxplots; dots are outliers.

or contaminating debris in subsequent analyses, we elected to exclude all 9 from further study. This procedure resulted in a final dataset of 21 high-quality single cell samples.

On average, 4.4 million paired-end fragments were obtained for each sample (Table B.4 in Appendix B), 2.7 of which could be mapped to annotated genes. These cover a mean of 4,717 ± 175.8 (SEM) genes per single cell; collectively, 13,582 different genes were expressed in at least one cell, which represents 74.2% of the genes expressed in the GFP^{high} OSNs. To confirm that the sequenced cells indeed correspond to GFP^{high} OSNs, I compared the transcriptome of each single cell to the expression profile of the WOM, the pooled 10 million OSNs and the bulk GFP^{low} and GFP^{high} populations. All single cells correlated better with the OSN samples than with the WOM (p-value < 2.2e-16, paired t-test), and better with the GFP^{high} than with the GFP^{low}(p-value < 2.2e-16, paired t-test; Figure 3.8).

Next, I examined the expression of a set of genes that are canonical markers for the different cell types found in the WOM and the VNO, the other tissue where OMP^{+} cells are abundant. All 21 cells showed robust expression of the genes characteristic of mature OSNs (such as *Omp*, *Gnal*, *Cnga2*, *Ano2* and *Adcy3*), and low or no expression of markers of other cell types, including immature cells and VSNs. Also, none of the 21 cells expressed GC-D or TAAR genes (Figure 3.9). Thus, this confirms that the cells

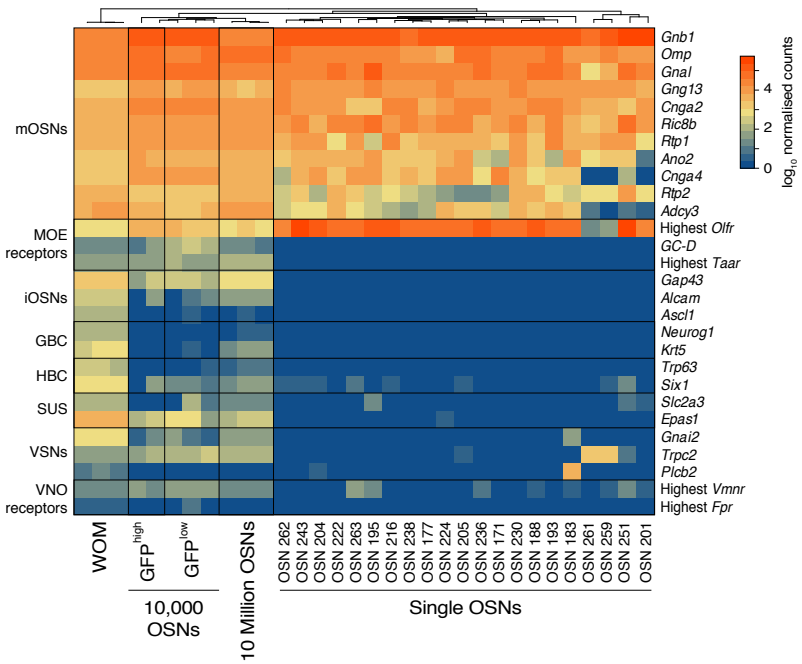


Figure 3.9 – **Expression of canonical markers.** Heatmap representing the expression levels for different genes that are canonical markers for various cell types found in the WOM and VNO. Single cells express all the marker genes of mature OSNs and virtually none of the genes from other cell types. mOSNs, mature OSN; iOSNs, immature OSN; GBC, globose basal cell; HBC, horizontal basal cell; SUS, sustentacular cell.

sequenced correspond to mature OSNs, with no contamination from other cell types.

3.3.1 Heterogeneity between single OSNs.

Single-cell RNAseq data suffers from much higher technical variation than bulk RNAseq, because the amount of starting material is very low. Many genes can have very different normalised counts between technical replicates taken from the same pool of total RNA. For example, a gene can have a hundred to a thousand counts in one replicate and zero in the other; only genes that are expressed at high levels are consistent[318]. Therefore, the variation observed between two different single cells will be a combination of the large technical noise plus true biological variation. Consistent with this, the correlation between any two single OSNs was only moderate ($0.45 < \text{rho} < 0.61$, p-value $< 2.2\text{e-}16$), and there was great variation in the expression levels of low and moderately expressed genes (Figure 3.10).

To address the question of how heterogenous are the transcriptomes of OSNs expressing different OR genes, I calculated the coefficient of variation (CV) for all those genes with more than a thousand normalised counts in at least one single OSN. At this expression level, variation should largely reflect biological variability instead of technical noise.

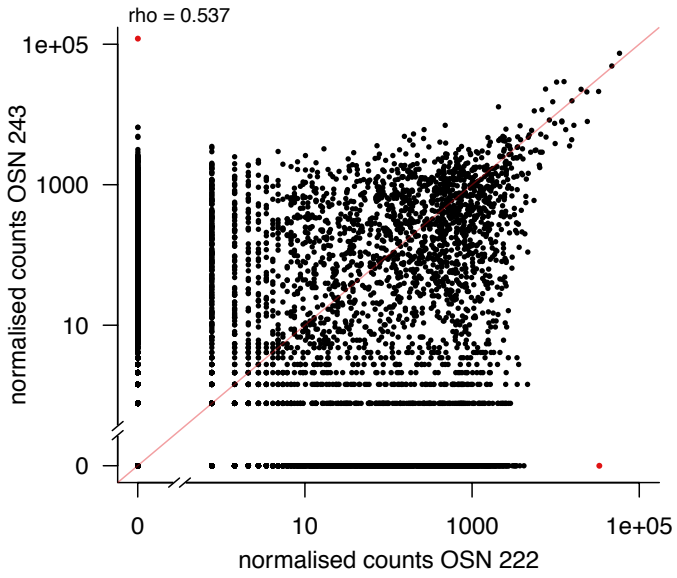


Figure 3.10 – Correlation between two single OSNs. Scatter plot of the normalised counts of two different single OSNs; 0.1 has been added to the counts before computing the logarithm, to be able to show the genes not expressed in one sample. The red line corresponds to the 1:1 diagonal. The Spearman rank correlation coefficient is indicated in the top left corner. In red are higlighted the two OR genes abundantly expressed by each OSN.

From these, 598 showed highly variable expression across cells (CV > 4; Figure 3.11A). These were relatively evenly distributed across the single OSNs (Figure 3.11B) except for one unusually variable cell (OSN 183) which contained 15% of all the genes; these were enriched in GO terms related to chemokine receptor binding, cytokine binding and activity, antigen processing and presentation and regulation of lymphocyte activation, which suggest a stressed cellular state. In contrast, the remaining 509 genes are only enriched in G-protein coupled receptor signalling and transduction. The significance of this term stems from the variable expression of OR genes, but there are also a few orphan GPCRs (*Gpr32*, *Gpr123*, *Gpr125* and *Gpr160*). A similar enrichment analysis on protein domains identified the expected seven-transmembrane receptor domain present in all ORs, but also a significant enrichment for a zinc-finger motif (C2H2 type) and a KRAB box domain, both found within a group of zinc finger proteins (ZFP). Thus, different OSNs are distinguished by the OR gene they express; ZFP genes might also pattern different subtypes of OSNs, but a larger sample is necessary to assess the significance of this finding and whether their expression is correlated with distinct subtypes of OSNs.

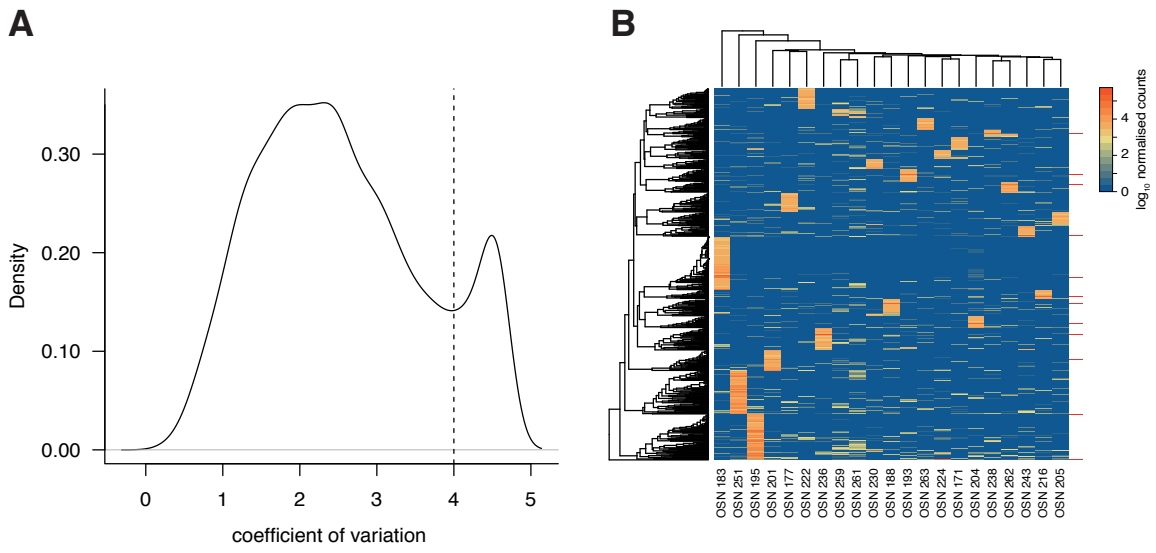


Figure 3.11 – Highly variable genes pattern single OSNs. A) Density distribution of the coefficient of variation (CV) for all those genes that have at least 1,000 normalised counts in at least one cell. **B)** The genes with a CV > 4 are highly variable between cells and only expressed in one or a few samples. To the right of the heatmap, a red line indicates the gene is an OR.

3.3.2 Monogenic expression of OR genes.

The expression of OR genes in OSNs of the MOE is considered to be monogenic. Many experiments have supported this paradigm, but it has never been conclusively proven[74]. Most evidence has come from both single and double *in situ* hybridisation, sometimes in OSNs expressing a receptor tagged with a reporter protein, to show that a given neurone does not express two receptors at the same time. But in all cases only a (very) restricted set of receptors has been tested; the lack of coexpression between these has then been extrapolated to the complete repertoire. Also, in a scenario where the coexpression of a given OR gene with any other is random, only a very small number of OSNs would be expected to have a particular combination of two receptors and, therefore, it would be very challenging to detect by traditional methods. Furthermore, since *in situ* hybridisation requires specific probes, many times it is not possible to distinguish between closely related receptors. RNAseq of single cells allows, for the first time, to investigate the complete receptor gene repertoire in each cell, and assess how many different ORs are expressed.

First, I analysed the expression of all OR and TAAR genes and pseudogenes, and identified 476 receptor genes with at least one fragment mapped in at least one single OSN; all of these were OR and no expression of TAAR genes was detected. However, the great majority of these OR genes (86%) had less than one third of their gene

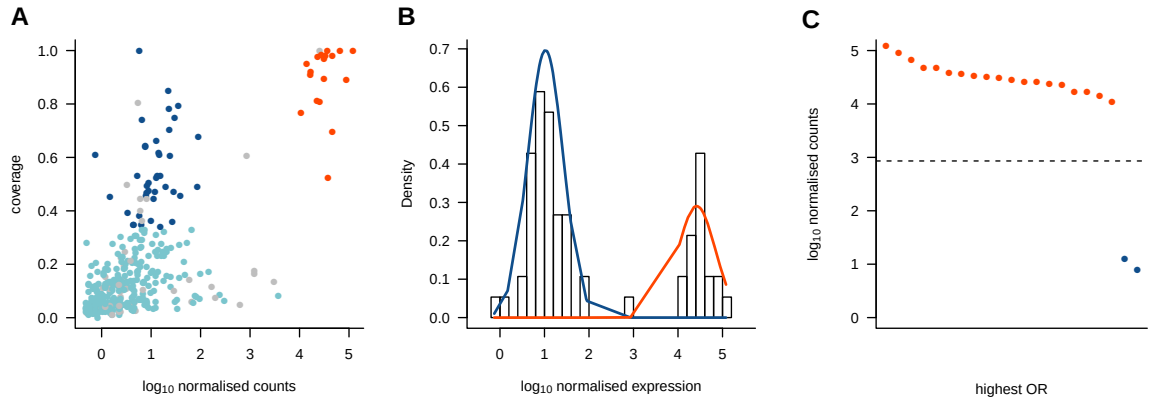


Figure 3.12 – OR expression in single OSNs. A) Scatter plot of the normalised expression of each OR gene in each single OSN against the proportion of the gene covered by at least one read. Pseudogenes are in grey. The majority of the OR genes are expressed at very low levels and have low coverage (light blue). The other 15% have coverage $> 1/3$ and **B)** are bimodally distributed into high (red) and low (dark blue) expressed OR genes. Two normal-like distributions were fit to the data to separate the genes based on their expression levels. **C)** Normalised counts for the highest OR expressed in each single OSN. In 19 of the 21 cells an OR is expressed at very high levels (red), while two OSNs show very low OR expression (dark blue). The horizontal dotted line indicates the expression value where the two normal distributions from (B) intersect, and represents the threshold to separate the high- from the low-expressers.

length covered by sequencing fragments, which suggests non-specific transcription and/or mismapped reads (Figure 3.12A). The remaining 65 OR genes segregated into two distinct distributions, depending on their expression level (Figure 3.12B). A total of 45 OR genes were expressed at low levels, with mean expression of only 15.9±2.7 (SEM) normalised counts, while the other 20 were expressed on average at 36,162.5±6,238.7 (SEM) normalised counts. The latter had sequencing fragments mapped along the majority of the full transcript, with median coverage of 0.93.

Next, I looked at the OR genes expressed in each single OSN. The intersection between the two distributions in Figure 3.12B (855.98 normalised counts) can be used to define whether an OR is expressed at low or high levels. In 17 of the 21 single OSNs there was a single OR gene, annotated as functional, expressed at high levels. Two more cells expressed an annotated OR pseudogene at similarly high levels*: Olfr1191-ps1* and *Olfr1224-ps1*. Closer inspection of these two genes, revealed that they encode full-length ORFs of 318 and 311 amino acids each, and these align to other mouse OR genes with high identity; *Olfr1224-ps1* is annotated as protein-coding in Ensembl. Thus, both of these genes are likely to encode a functional receptor. Together, 19 of the 21 single OSNs express a single putatively functional OR gene at great abundance (Figure 3.12C). Indeed, the OR genes rank on average as the 6^{th} most abundantly expressed gene in the transcriptome of these OSNs; only *Stoml3*, *Gnb1*, *Malat1* and *Calm1* are consistently expressed at higher levels. All the expressed OR genes are Class II except for *Olfr556*.

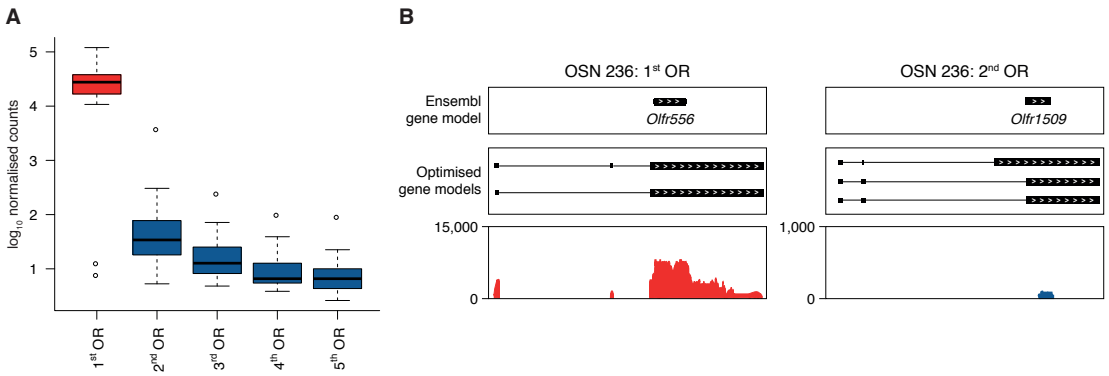


Figure 3.13 – Monogenic expression of OR genes. **A)** Boxplots for the five most abundant putatively functional OR receptor genes expressed in each single OSN, in decreasing order. Most OSNs express one OR gene at very high levels (red). The next most abundant OR gene is, on average, a thousand times lower. **B)** Representative example of the first and second most abundant OR genes expressed in a particular OSN. On the top panel is the gene model present in Ensembl, with the reconstructed models from the RNAseq data below. Each black box represents an exon and the lines joining them are introns; the arrow heads indicate the strand of the gene. At the bottom is a coverage plot of the sequencing reads. For the most highly expressed receptor (red) a large number of sequencing reads cover the full gene model, but for the second most abundant OR gene (blue) there is only a small number of reads that do not span the whole gene. Note the difference in scales for each plot.

In addition to the single abundant functional OR, the 19 single OSNs had between 11 and 28 other OR genes with evidence of expression, but all had extremely low normalised counts. After excluding annotated pseudogenes, the most highly expressed OR gene is on average over 1,000 times more abundant than the next highest OR gene expressed (Figure 3.13A-B). In two cases, a pair of single cells expressed the same OR gene; OSN 171 and OSN 177 each express *Olfr728*, and OSN 22 and OSN 263 each express *Olfr55*. The OR genes expressed at low levels in these two pairs of cells are not shared to a greater degree as do any other two cells, thus suggesting that these lowly expressed genes are not coordinated with the expression of the abundant OR gene, nor are they the product of mismapping.

To assess whether these low levels of OR expression could be biologically meaningful or whether they are more likely to represent leaky expression, I analysed publicly available single cell RNAseq data from 96 mouse T-helper lymphocytes[319] and 288 single mouse embryonic stem cells[320], which were captured and processed in a similar manner as the single OSNs, in the same facility. In these datasets, some cells expressed up to 101 OR genes, but all at very low levels (Figure 3.14). Therefore, this suggests that the low expression of a fraction of the OR repertoire is mainly the result of non-specific transcription and has no biological significance. The contrast with the highly expressed receptor, along with the recapitulation of this pattern in non-olfactory cells, strongly argue in favour of a monogenic expression pattern of OR genes in OSNs.

As discussed in Chapter 2, many OR genes have several transcripts that differ in their

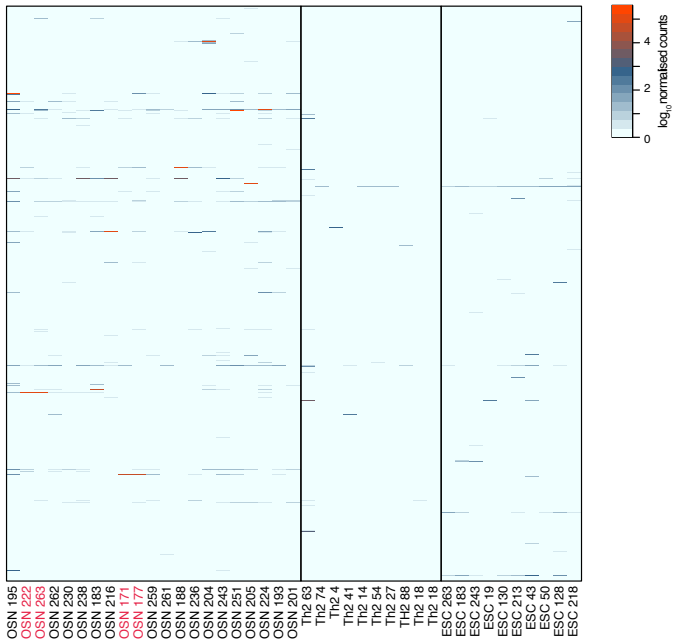


Figure 3.14 – OR expression in several single-cell datasets. Heatmap of the normalised expression levels for the complete OR repertoire; each row represents a different receptor gene. All 21 single cells are presented alongside 10 representative T-helper lymphocytes (Th2) and 10 embryonic stem cells (ESC). In the non-olfactory cells, there is a variable number of OR genes with low normalised counts, similar to what is observed for OSNs. The two pairs of single OSNs that express the same abundant OR gene are highlighted in red.

UTR structure and a few also have potentially different protein coding isoforms. None of the OR genes expressed in the single OSNs include genes with alternative protein isoforms, but most of them have several different transcripts. For these, the sequencing reads support expression of several distinct transcripts with alternative UTRs. This demonstrates that within a single OSN, the different isoforms of the OR gene expressed abundantly are present.

3.3.3 Monoallelic expression of OR genes.

The OMP-GFP mouse line is in a mixed genetic background of 129P2×C56BL/6 strains [309]. This allows to discriminate which allele is expressed, for all those OR genes with SNPs between the two strains. Therefore, to assess whether the abundant OR gene is expressed in a monoallelic fashion, I mined the mouse genomes project (MGP) data[321] to obtain the variable nucleotide positions within the highly expressed ORs. Of the 19 abundant OR genes, nine had at least one SNP within exons. As an example, *Olfr55* is expressed in both OSN 222 and OSN 263, and has 15 SNPs across its transcript. For OSN 222, there are 54,677 reads that map across these, and 54,563(99.79%) support the

nucleotides found in the C57BL/6 genome. Similarly, there are 52,503 reads spanning variable positions in OSN 263, 52,456 (99.93%) of which have the variant of the 129P2 allele (Figure 3.15A). Since both these cells originated from the same mouse, these data directly demonstrate that *Olfr55* is expressed monoallelically. For the other seven genes with SNPs, five expressed exclusively the C57BL/6 allele and two expressed the 129P2 allele; in all these cases, at least 99.57% of the reads supported expression of one of the alleles (Figure 3.15B). Thus, I have directly demonstrated that OR expression in single OSNs, not only is monogenic, but it conforms to an extremely tight monoallelic expression pattern.

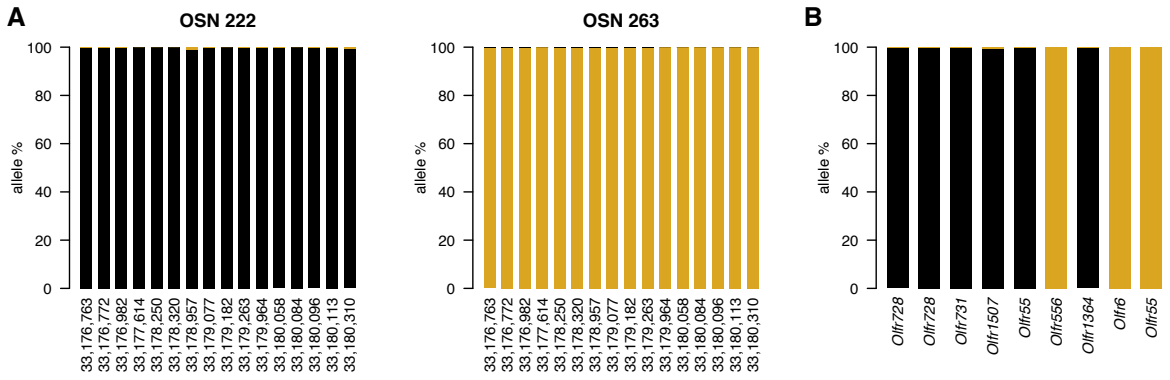


Figure 3.15 – OR expression is monoallelic. A) Stacked barplots of the proportion of sequencing reads supporting the C57BL/6 (black) or 129P2 (golden) alleles for the *Olfr55* gene in OSN 222 and OSN 263. Each bar represents a different SNP. **B)** Same but for the cummulative data for all SNPs in each of the other OR genes expressed at high levels that have SNPs. In all cases, a single allele is supported by over 99.5% of the data, indicating very tight monoallelic expression.

3.3.4 Identification of a novel type of OSN.

Two of the sequenced single OSNs (OSN 259 and OSN 261) did not express any annotated OR gene at high levels (Figure 3.12C); they have some expression of ORs but all are very low (the highest are at 12.66 and 7.7 normalised counts respectively) and rank at similar levels to what is found in the T-helper lymphocyte and ES cell transcriptomes. I could not detect expression of other chemoreceptor genes either, such as vomeronasal or taste receptors. From Figure 3.9 it is evident that neither of these cells express *Adcy3* or *Cnga4*, both essential components of the canonical signal transduction pathway, but they express many other markers of mature OSNs such as *Omp*, *Gnal* and *Cnga2*. In contrast, they express *Trpc2*, a gene characteristic of VSNs. During the course of this PhD, two subpopulations of OSNs in the MOE were shown to express *Trpc2*[131].

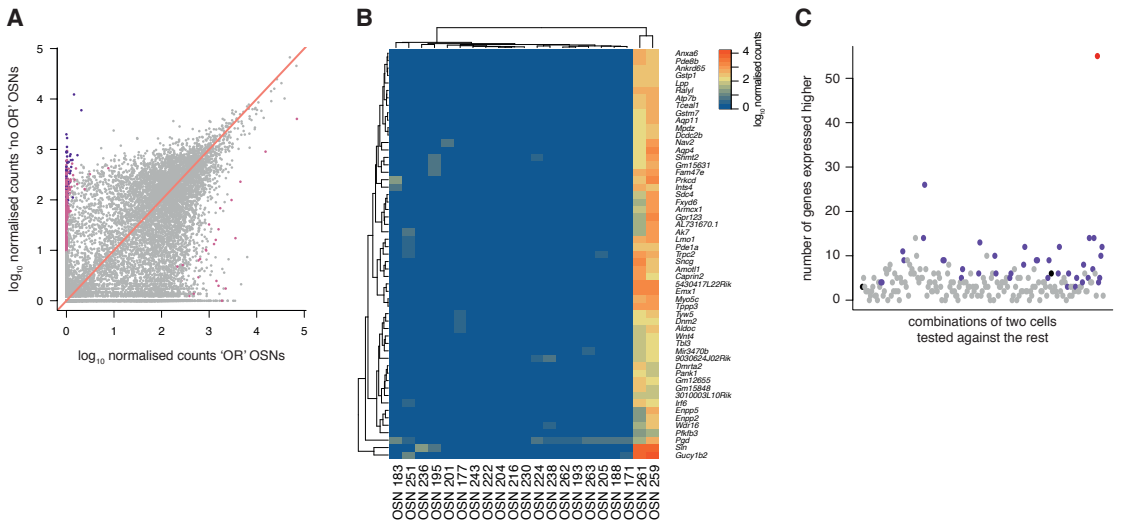


Figure 3.16 – Characteristic expression profile of a novel type of OSN. A) Scatter plot of mean normalised expression values for the whole transcriptome in the 19 OR-expressing cells (x-axis) versus the two no-OR cells (y-axis). The red line is the 1:1 diagonal. Genes significantly DE (FDR < 5%) are in pink, and those consistently expressed in both no-OR cells are in purple. **B)** Heatmap of the 55 DE genes that are consistently expressed higher in the no-OR cells compared to the other 19 cells. **C)** Number of genes that are consistently expressed higher when two cells are compared to the other 19. Each dot represents a specific combination of two samples. In purple are all those combinations that include one of the no-OR cells, and in red the combination of both no-OR cells.

In order to characterise these cells better, I compared their transcriptomes to those of the other 19 –OR-expressing– cells. Differential expression analysis revealed 494 significantly DE genes (Figure 3.16A) but only 55 that were consistently expressed in both cells (Figure 3.16B). In order to determine if this number of shared DE genes is statistically meaningful, I performed the same analysis for all 210 possible combinations of two among the 21 single OSNs. This pair of 'no OR' cells have over twice as many DE genes as any other pair of cells (Figure 3.16C) which suggests that they are indeed different from the rest and posses a set of genes specifically expressed in this novel type of OSNs.

The most abundant DE gene is *Gucy1b2*, a soluble guanylyl cyclase, followed by sarcolipin (*Sln*) and *Emx1*, a transcription factor involved in neuronal fate specification. *Trpc2* is ranked as the 39^{th} most abundant gene in the transcriptomes of these cells. In collaboration, Luis Saraiva and Masayo Omura validated the expression of some of the DE genes by *in situ* hybridisation, using *Trpc2* as a marker of this type of OSN. They found that *Gucy1b2* and *Sln* indeed are expressed in the MOE, sparsely distributed within the OSN and sustentacular cell layers (Fig 3.17A-B). By two-colour *in situ* they could confirm that the *Gucy1b2*$^+$ cells define a subset of the *Trpc2*$^+$ cells in the MOE (Fig 3.17C), consistent with the single-cell RNAseq data and recent reports[131, 322]. Additionally, *Gucy1b2* was found to be coexpressed with *Sln* and *Sncg* (a less abundant

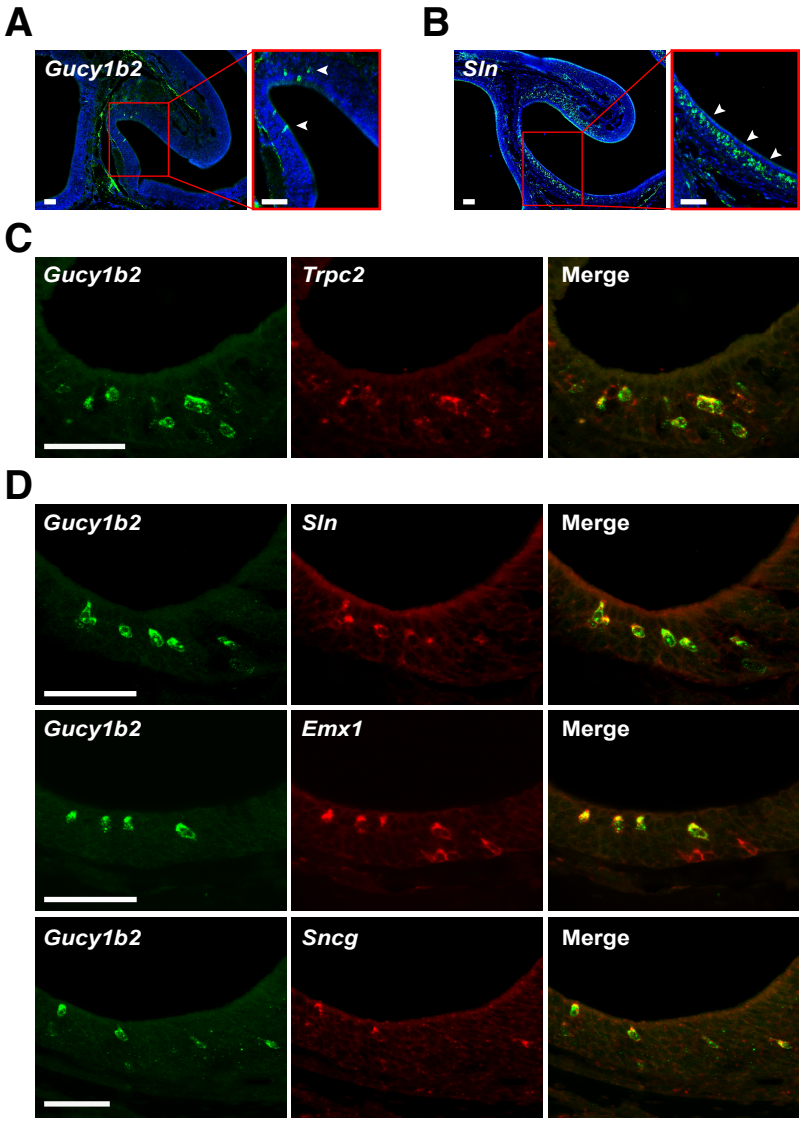


Figure 3.17 – Validation of the DE genes in no-OR cells. A-B) Cryosections of adult mouse MOE hybridised with cRNA probes for the two most highly expressed DE genes *Gucy1b2* (A) and *Sln* (B). The hybridisation signals are sparsely distributed within the MOE. **C)** Two-color *in situ* hybridisation of the top ranked marker (*Gucy1b2*) with *Trpc2*. Some but not all of the *Trpc2*$^+$ cells in the MOE also coexpress *Gucy1b2*. **D)** Two-color *in situ* hybridisation of the top ranked marker (*Gucy1b2*) with other differentially expressed genes, *Sln*, *Emx1* and *Sncg*. Arrowheads point to labelled cells. Scale bars, 50 μm. Image kindly provided by Luis Saraiva and Masayo Omura.

DE gene, ranked 7^{th}) and to partially overlap with cells expressing *Emx1* (Fig 3.17D). Thus, these two cells appear to be examples of the recently discovered type B *Trpc2*$^+$ cells in the MOE, which are now characterised by at least a handful of other genes, and clearly represent a distinct subtype of OSN.

In all, RNAseq has demonstrated to be a very powerful technology to definitely answer questions that are fundamental for the understanding of the olfactory system. The

data presented here strongly supports the monogenic character of OR expression, but a larger sample size will be necessary to extend this conclusion to more OSN types; the inherent challenges of working with single OSNs, however, will make this a difficult task. Interestingly, the monoallelic character of OR expression is extremely tight, with nearly the entirety of the sequencing data supporting expression of a single allele. Therefore, it seems like the expression of one abundant OR allele is very tightly controlled, with only a small degree of leaky transcription from a small subset of OR genes.

Chapter 4

Genetic variation and the expression of the OR repertoire.

I have demonstrated that high-depth RNAseq is a suitable technology to study the transcriptome of the olfactory system, and that the expression of the majority of the receptor repertoire can be detected, albeit at low levels. The amount of sequencing data devoted to the receptor genes can be increased by separating the OSNs from the other cell types found in the tissue, but the increase is only modest (2.5 fold). These experiments revealed that the profile obtained from the whole tissue is virtually identical to that from sorted OSNs, indicating that despite the lower amount of data obtained from WOM samples, the information is accurate and reproducible. The advantage of analysing the transcriptome of the WOM is that wild-type animals can be used, without the need for any transgenes. Therefore, I decided to continue to profile WOM samples. Also, since the OR genes are more divergent than VRs, these can be accurately detected by unique counts, and thus I focused my subsequent analyses on the OR repertoire only.

The data presented so far has revealed a characteristic expression distribution for the OR repertoire; a few genes are highly abundant and expression values then drop quickly. Remarkably, each of several animals from the same genetic background contains a stereotyped proportion of each receptor type, suggesting that OR gene choice is very tightly regulated. But how rigid is this pattern of expression and what are the factors contributing to such a stable state?

4.1 Gender has little effect on OR gene expression.

The C57BL/6J (hereinafter B6) WOM transcriptome data presented in Chapter 2 was obtained from adult male and female samples. Among the behavioural responses elicited through olfactory signals, many are clearly distinct between adult male and female mice, including sexual conduct[157, 159, 323], aggressive responses to intruders[156] and parental care[128], but the mechanisms that ensure such differentiated responses have not been fully elucidated in mammals[324]. It is conceivable that the receptors involved in the detection of the chemicals mediating these dimorphic behaviours might be differentially expressed between sexes. To assess this, I performed a differential expression analysis. Surprisingly, the transcriptomes showed a striking similarity between males and females (Figure 4.1). At the whole genome level, only 32 genes were significantly DE (FDR $< 5\%$). Among these are those expected to be different by sex, such as genes from the Y chromosome or the X-inactive specific transcript, *Xist*. Except for these, the fold-changes observed were small, with only a mean 1.56 difference between the sexes, for both genes expressed higher in males and in females. The expression estimates for each OR gene between males and females were also remarkably similar, with only 0.7% of the repertoire significantly DE (9 OR genes; Figure 4.2). This suggests that the observed behavioural dimorphism is not achieved by a differential ability to detect certain olfactory cues.

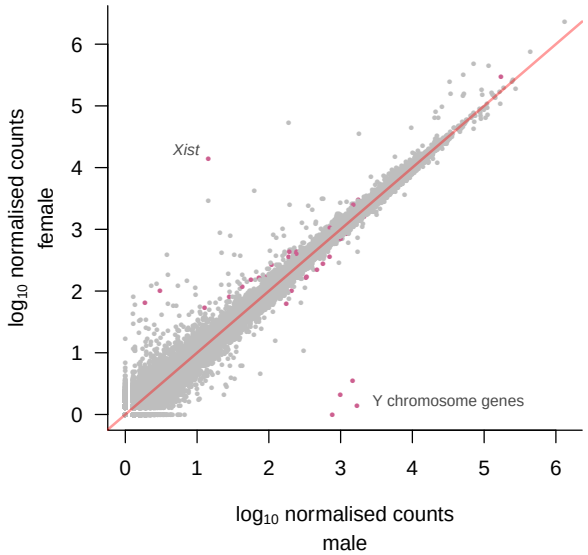


Figure 4.1 – Transcriptome of the WOM of males and females. Scatter plot of the mean normalised counts for the transcriptome of the WOM in males and females. The red line indicates the 1:1 diagonal. Most genes lie close to the diagonal, suggesting that the transcriptomes are nearly equivalent. The 32 significantly differentially expressed genes are highlighted in pink; among these are genes known to be expressed only in one sex, such as *Xist* or genes from the Y chromosome.

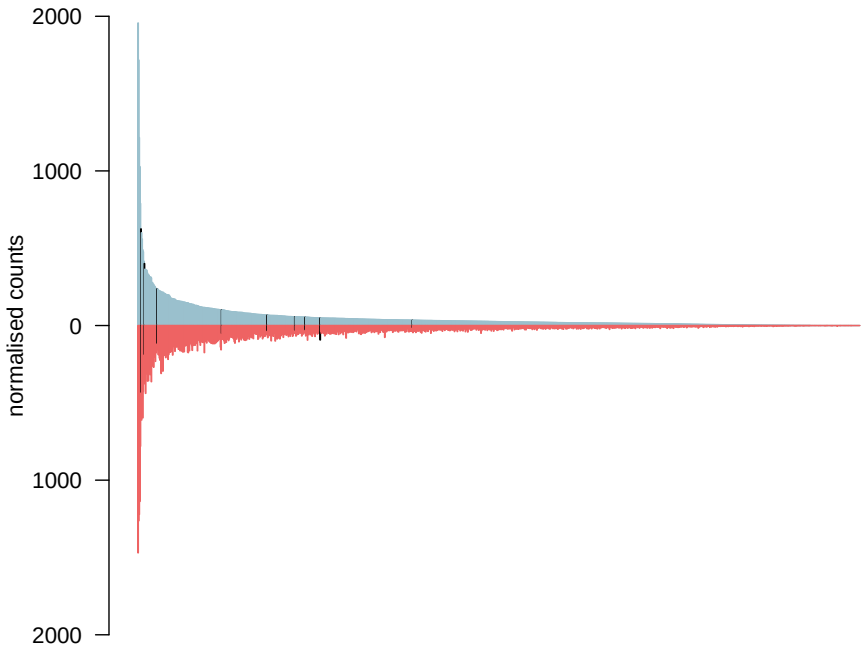


Figure 4.2 – OR expression in males and females. Barplot of the normalised expression of the OR genes in males (blue) versus females (red). Counts have been normalised for both depth of sequencing and gene length. Genes are ordered by decreasing male expression, and the corresponding values in females are presented as a mirror image. The repertoires are extremely similar. Significantly differentially expressed genes are highlighted in black.

4.2 Some OSN types are more abundant than others.

The observed distribution of expression values for the OR genes is conserved between individuals and is not significantly affected by sex. Such an expression profile may be explained by two scenarios: either 1) OR genes with high expression values are expressed in a larger population of neurones than those with low expression values; and/or 2) OR genes are consistently expressed at different levels per cell. To differentiate between these possibilities, I first compared the RNAseq expression values with cell counts in the MOE, for a subset of OR genes. These data were obtained by counting fluorescent OSNs from transgenic mice (21 days of age) carrying tagged receptors[325]. Since all these transgenic lines are in a mixed 129×B6 genetic background similar to the OMP-GFP line, I compared to the RNAseq data from the OMP-GFP WOM (25 days of age) from Chapter 3. The correlation with the RNAseq normalised counts was very strong and highly significant (rho = 0.94, p-value = 0.00047; Figure 4.3A). Thus, the expression values obtained through RNAseq are a reflection of the number of cells expressing a particular receptor gene within the MOE.

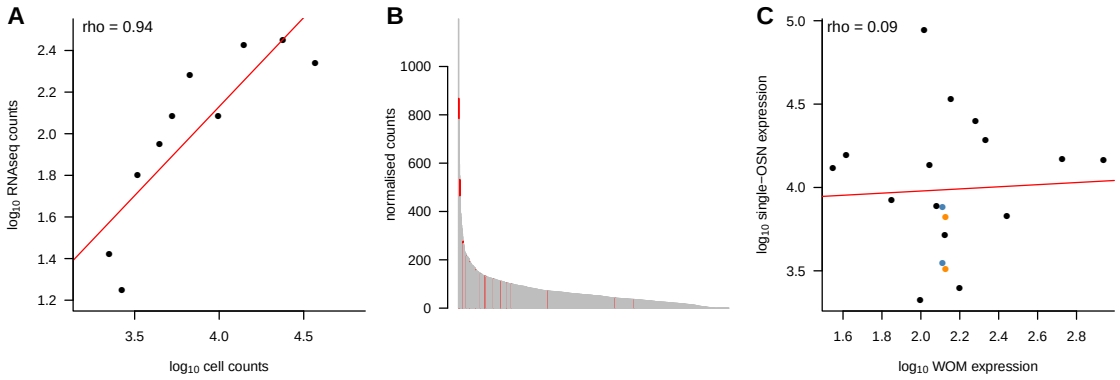


Figure 4.3 – RNAseq expression correlates with neurone number A) The expression levels for a subset of OR genes are plotted against their count number in the MOE. In red is the regression line and the Spearman correlation coefficient is on the top left corner. The high correlation indicates that RNAseq expression levels reflect the number of cells expressing a particular OR gene in the MOE. **B)** Barplot of the expression of the OR repertoire in the OMP-GFP WOM data, in decreasing abundance order. Those receptors that were expressed in the 19 single OSNs sequenced are highlighted in red. These cluster towards the highly expressed genes which suggests that the MOE has higher number of OSNs expressing these receptor genes, which results in a higher probability of sampling them. **C)** The RNA expression in the OMP-GFP WOM data is plotted against the expression levels of abundant OR genes in single OSNs. The two single cells expressing *Olfr55* are in orange and the two expressing *Olfr728* are in blue. In red is the regression line. There is no correlation between the two, suggesting that the cumulative expression seen at the population level is not influenced by the OR abundance per OSN.

These data indicate that the first scenario might be the major contributor to the observed expression profile for the receptor repertoire. If this is true, one would expect that the OR genes expressed in randomly sampled neurones from the MOE would be enriched for those expressed at high levels. Such a random sampling of OSNs was performed when producing the single-cell RNAseq data presented in Chapter 3. Indeed, from the 17 different OR genes expressed at high levels in the 19 single OSNs, 14 are within the top quartile of the distribution and this enrichment is highly significant (p-value = 9.23e-07, hypergeometric test; Figure 4.3B). Thus, there is a strong bias towards selecting receptors with high RNAseq expression estimates within the WOM. To test if there is also a contribution from varying levels of OR expression per OSN, I compared the expression levels in the WOM RNAseq data against the expression of each OR in the individual OSNs; to account for variability, I normalised to five OSN marker genes (*Omp*, *Gnal*, *Adcy3*, *Ano2* and *Cnga2*) that have been shown to be stably expressed[304]. While the OR gene expression levels did vary within single cells (mean = 15,353.57 and SD = 19,483.62 normalised counts), they did not correlate with the corresponding expression levels in the WOM (rho = 0.09 p-value = 0.71; Figure 4.3C). Together, these data suggest that while there is variance in the OR expression levels between different OSNs, this does not correlate to their cumulative abundance in the overall population. Instead, the expression levels obtained from WOM samples reflect the number of cells expressing a particular OR gene. Therefore, RNAseq is an accurate proxy for quantifying

the diversity of OSN types found within the MOE, such that high RNAseq expression equates to high number of OSNs.

4.3 OR expression differs between mouse strains.

The relative proportion of each OSN type is stable between genetically identical animals irrespective of sex (Figure 4.2). To investigate whether this OSN distribution is a feature of all mice, I analysed the WOM transcriptome of a different laboratory strain, 129S5SvEv (referred to as 129; Tables B.1 and B.2 in Appendix B). The 129 genome has 4.4 million single nucleotide polymorphisms (SNPs) and 810 thousand small indels (1-100 bp long) compared to B6[321], of which 13,484 SNPs and 1,936 indels are found within OR gene transcripts. All these variants can have an impact on the mapping of the RNAseq data, especially since ORs are more variable than the average gene[271]. To assess this, I mapped the data to the B6 reference genome, or to a pseudo-129 genome that contains all the high quality SNPs and short indels reported for this strain[321]. Despite the great majority of the repertoire having similar expression estimates in both cases (85.6% have changes of less than one normalised count), a few receptor genes show a significant change in the number of sequencing fragments that map to them (Figure 4.4). In most of these cases (84.4%), the pseudo-129 genome allows a greater proportion of reads to be mapped. Thus, I utilised the data mapped to the pseudo-129 genome for further analyses.

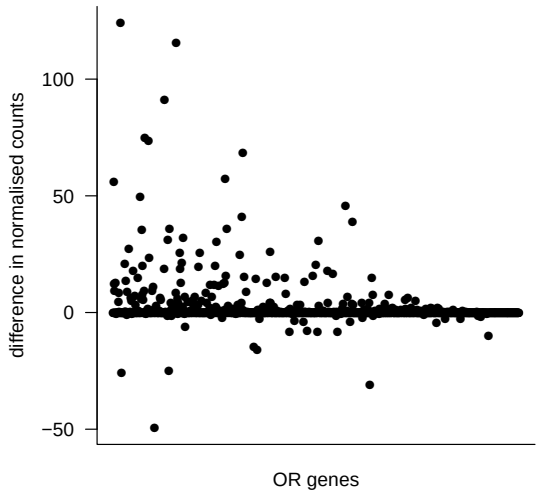


Figure 4.4 – Effect of imputing variation on OR expression estimates. Plot of the difference of the mean expression values for OR genes, as obtained by mapping to a pseudo-129 genome versus mapping to the B6 reference. The genes are ordered by their decreasing mean expression value in the calculations using the pseudo-129 genome.

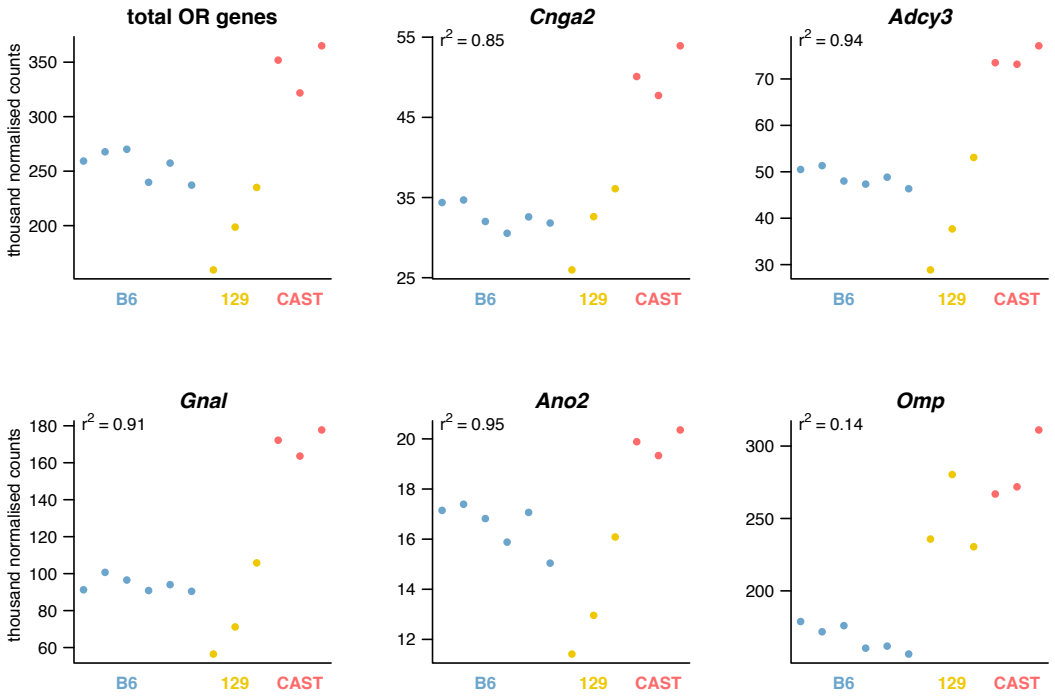


Figure 4.5 – Normalisation for OSN number. In the first panel the total normalised counts for OR genes are plotted for each WOM sample from different strains; CAST data will be presented below. Then, the normalised counts are plotted for a set of marker genes that are known to be stably expressed in OSNs only: *Cnga2*, *Adcy3*, *Gnal*, *Ano2* and *Omp*. The Pearson's correlation value with the total counts is indicated. All marker genes except for *Omp* are highly correlated to the total number of counts in OR genes, suggesting that the observed differences are the product of different proportions of OSNs in the WOM samples.

To be able to compare the relative distribution of OR gene expression in mature OSNs across different strains, it is necessary to normalise to account for any differences in the total number of neurones present in WOM samples. For this, I used a method proposed by Khan et al.[304]. This approach uses marker genes known to be stably expressed in mature OSNs only, to estimate the proportion of WOM RNA contributed by the OSNs. As can be observed in Figure 4.5, the total number of normalised counts across the receptor repertoire varies considerably between strains, even after normalising for differences in depth of sequencing. Inspection of genes only expressed in OSNs (*Cnga2*, *Adcy3*, *Gnal*, *Ano2* and *Omp*) revealed that most of them follow the same pattern of expression and are highly correlated with the total OR counts (Figure 4.5). Therefore, the observed differences are most likely due to different proportions of OSNs in the WOM samples. In this particular case, *Omp* had a very poor correlation value ($r^2 = 0.14$, p-value = 0.22, Pearson) and thus was not used in the normalisation process. To normalise for OSN number, size factors were obtained based on the geometric mean of the OSN markers, and these were used to scale the OR genes normalised counts.

To evaluate how similar are the OR expression distributions between the animals

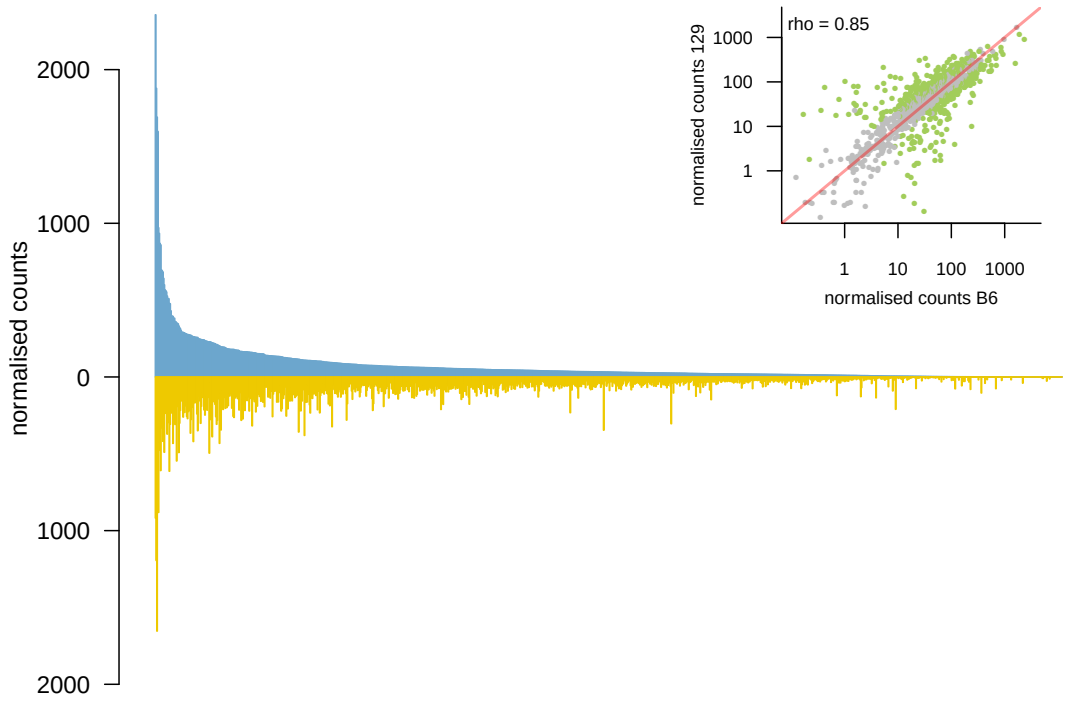


Figure 4.6 – OR expression in B6 and 129. Barplot of the OR normalised expression values in B6 (blue) with the corresponding 129 values (yellow) as a mirror image. As an inset is the scatter plot of the same expression levels; the red line is the 1:1 diagonal and the Spearman correlation coefficient is indicated. Significantly differentially expressed genes (FDR < 5%) are highlighted in green. Many OR genes have differences, albeit small, in their expression levels.

from these two genetic backgrounds, I performed a differential expression analysis on all OR genes. From these, 462 were significantly DE (FDR < 5%), which represents 37% of the whole repertoire (Figure 4.6). However, 45.9% of the DE genes had a difference lower than 2-fold, implying consistent but relatively small changes in expression. Concordant with this amount of DE genes, the correlation for OR gene expression was 0.85 (p-value < 2.2e-16, Spearman), that contrasts with the very high correlation of 0.96 (p-value < 2.2e-16, Spearman) across the whole transcriptome.

To determine whether genetic diversity influences the variance in OR gene expression, I repeated this experiment using CAST/EiJ, a wild-derived strain from the *Mus musculus castaneus* subspecies (henceforth referred to as CAST; Tables B.1 and B.2 in Appendix B). This strain has 17.6 million SNPs and 2.7 million short indels relative to B6; of these, 45,688 SNPs and 6,303 indels are found within OR transcripts. After mapping to a pseudo-CAST genome, 634 OR genes were significantly DE (FDR < 5%) compared to B6, constituting 50.8% of the whole OR repertoire (Figure 4.7). The changes in expression for some OR genes were dramatic: 132 genes had differences of at least 8-fold and therefore the OR gene correlation between the two strains was only 0.73 (p-value <

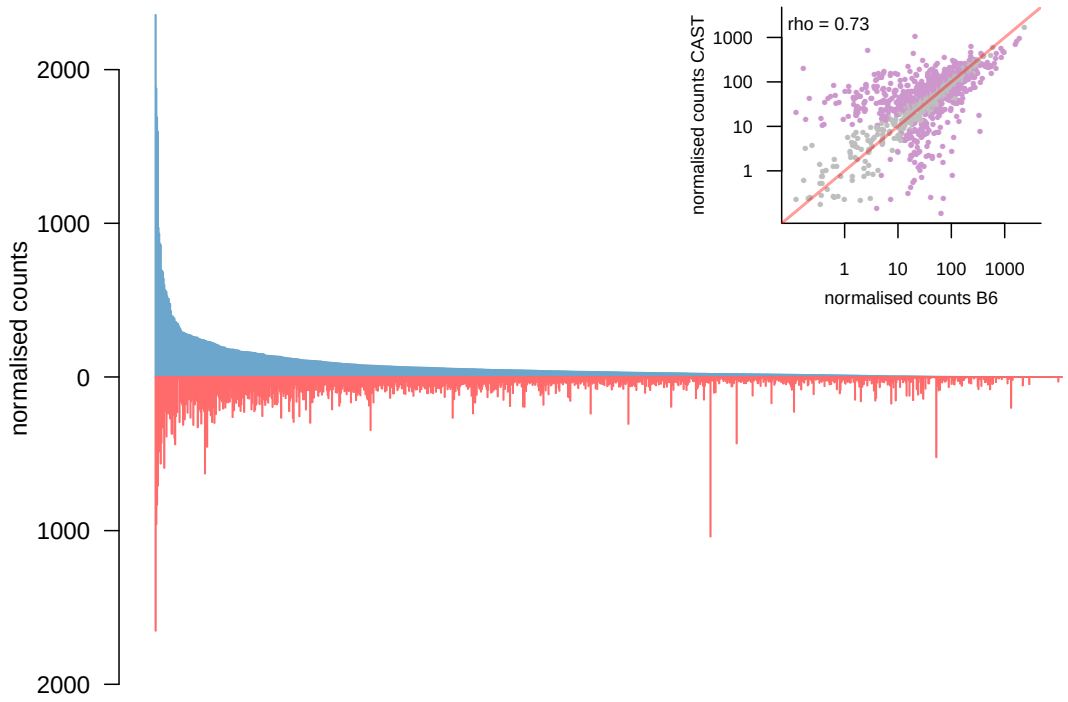


Figure 4.7 – OR expression in B6 and CAST. Barplot of the OR normalised expression values in B6 (blue) with the corresponding CAST values (red) as a mirror image. As an inset is the scatter plot of the same expression levels; the red line is the 1:1 diagonal and the Spearman correlation coefficient is indicated. Significantly differentially expressed genes (FDR < 5%) are highlighted in purple. Many OR genes have distinct expression levels and some are strikingly different.

2.2e-16, Spearman) while at the whole genome level the correlation remained high, at 0.96 (p-value < 2.2e-16, Spearman).

Taking all pairwise comparisons into account (including 129 vs CAST), 821 OR genes (65.7%) were DE between at least two strains (Figure 4.8A). Of these, 136 were DE in all three comparisons (Figure 4.8B), indicating a consistently different level of OR gene expression in each strain. All together, I have profiled the complete OR repertoire of three different strains of mice, with varying levels of genetic divergence among them. The data indicates that the OR expression levels are tightly controlled; a very stable distribution is conserved among animals of identical genetic background. However, the composition of the neuroepithelium of the WOM is remarkably diverse between animals with a different genetic makeup. Since the differences in expression levels indicate differences in the number of OSNs expressing particular OR genes, the MOE of each strain is a mosaic composed of varying proportions of each of the ~1,000 OSN types.

In order to investigate how the genetic background could be affecting OR expression levels, I mined the Mouse Genomes Project catalogue of SNPs and short indels for the 129 and CAST genomes[321]. Differentially expressed OR genes had greater amounts of

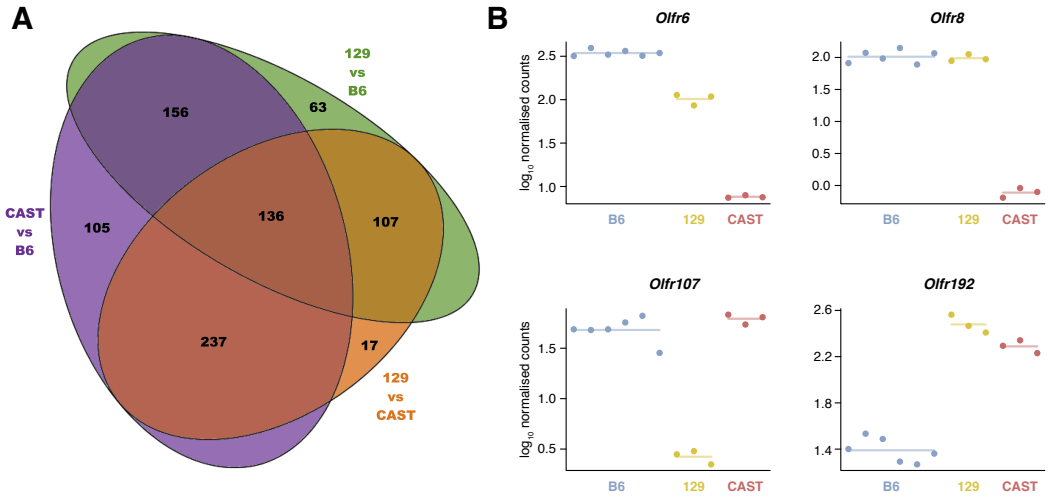


Figure 4.8 – Differential expression of the OR repertoire in three strains of mice. A) Proportional venn diagram indicating the number of OR genes that are significantly DE in the different pairwise combinations between the three strains. 136 receptor genes have different expression levels in all three strains. **B)** Examples of OR genes that are differentially expressed. Each point is a biological replicate and the horizontal line indicates the mean expression for each strain. In order, a receptor with differing expression values in each strain, one that is different in CAST only, in 129 alone or different in B6.

variation in their CDS, whole transcript or regions of 300bp or 1kb upstream of the TSS, in comparison to non-DE genes (Mann-Whitney, one tail), for both the 129 and CAST genomes (Figure 4.9). However, there was no relationship between the number of variable positions and the fold-change between the strains. The analysis of the putative promoters of many OR genes has shown that these have several OE-like and homeodomain binding sites[192–194], and the latter have been demonstrated to influence the frequency with which an OR gene is chosen for expression[188]. Therefore, it is likely that higher amounts of variation increase the probability of affecting some of these motifs, which would result in changes in the overall expression in the MOE.

Variation can also impact the open reading frame directly. For example, the B6 genome contains a premature stop codon in *Olfr421-ps1*, which results in a truncated protein of only 269 amino acids. On the other hand, the 129 and CAST genomes have a SNP that reverts the stop codon into a coding amino acid and, therefore, both produce a full-length protein (315 amino acids). OSNs that initially choose a pseudogene switch to another functional OR gene and progressively extinguish the expression of the non-functional receptor[207]. Consistent with this, the B6 animals have very low expression of *Olfr421-ps1*, while it is expressed more than ten fold higher in the 129 and CAST WOM (Figure 4.10).

Several examples exist on the influence of genetic variation on OR gene choice, such as the H-element which has been shown to elicit stronger effects the closer it is to the

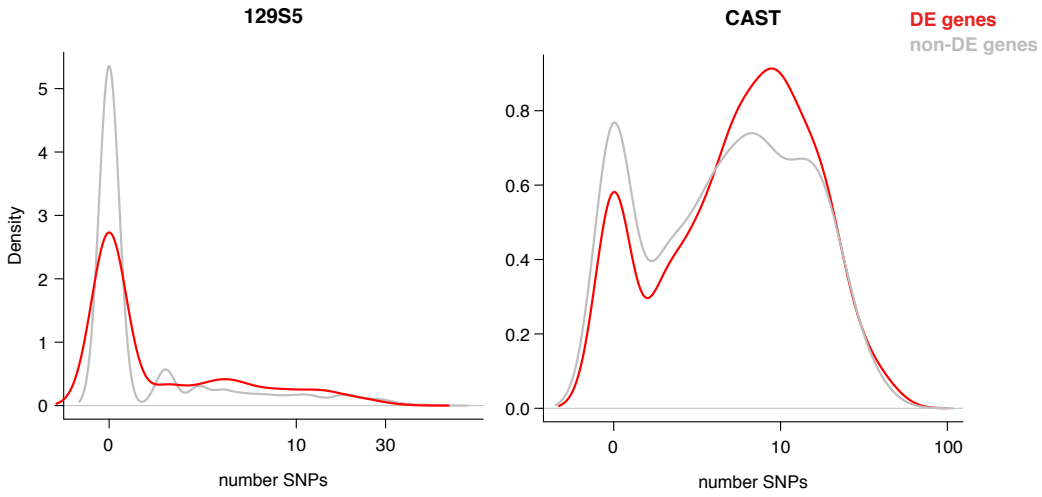


Figure 4.9 – Differentially expressed OR genes are more variable. Density distribution of the number of OR genes that have a given number of SNPs in the 1kb region upstream of the TSS. In grey are all those genes that are expressed at equivalent levels between B6 and 129 (left) or CAST (right); and in red, are the significantly differentially expressed genes (FDR $< 5\%$). There are less DE genes that have no SNPs and greater numbers of DE receptors have larger numbers of SNPs.

OR cluster it regulates[182]. The distance in the 129 genome is greater than in B6 animals[183] and, therefore, the expression levels of the proximal genes of the MOR28 cluster are lower in 129 than in B6. Enhancers, like the H-element and others, also contain transcription factor binding sites[188]. They can be located tens to hundreds of kilobases away from the OR genes they regulate. Commonly, they affect several OR genes

Figure 4.10 – Expression of a polymorphic pseudogene. *Olfr421-ps1* is an OR gene that in the B6 genome has a premature stop codon, but produces a full-length protein in the 129 and CAST animals. On the top is the gene model; the thicker box indicates the CDS and the rest is annotated UTRs. The blue line indicates the SNP that produces the STOP codon in B6. Below are representative examples of the raw sequencing data obtained from B6, 129 or CAST WOM. Each read is drawn in grey and the blue lines indicate the read spans exon-exon junctions; from the data it can be inferred that the gene model is lacking an additional 5' non-coding exon. Coloured lines within the reads indicate mismatches. The expression is many fold higher in the strains with a functional gene.

from the most proximal cluster[183, 186, 187]. Inspection of the pattern of differential expression in terms of the cluster organisation of OR genes revealed that, often, several adjacent genes are regulated in a concerted manner, with all genes expressed at higher or lower levels in a given strain. One such example is presented in Figure 4.11; a cluster of 8 OR genes, 6 of which are functional, consistently show lower expression in B6 than in 129 and CAST.

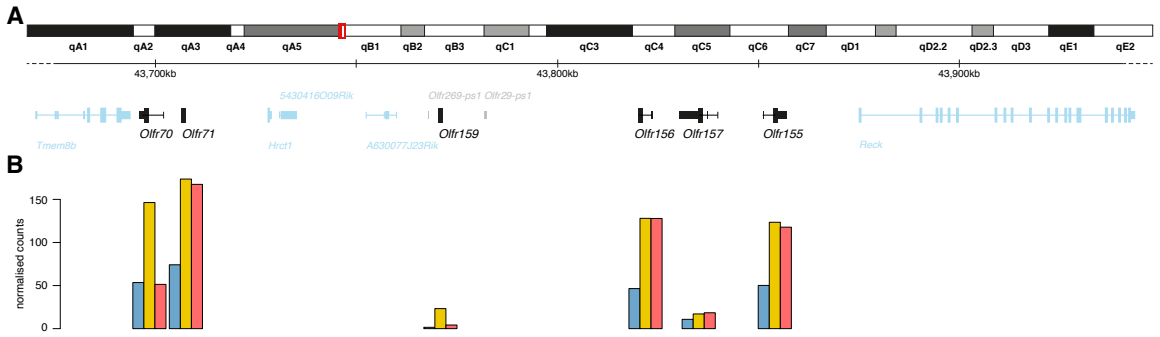


Figure 4.11 – Differentially expressed OR genes are clustered in the genome. A) Schematic representation of a cluster of OR genes. On the top is chromosome 4 and the region in the red box is expanded below. Each gene model is drawn with boxes for exons and connecting lines for introns. The height of the box indicates whether the sequence is protein-coding (taller) or not. A cluster of 8 OR genes spans 162.6kb; six of the genes are annotaed as functional (black) and two are pseudogenes (grey). Non-OR genes are shown in blue. **B)** The normalised expression levels for each of the functional OR genes is shown. B6 expression (blue) is lower than 129 (yellow) and CAST (red).

Thus, the differences in expression observed between the different strains of mice analysed here are likely to be the product of genomic variation located in important regulatory elements, both in the promoters of the OR genes and in long-acting enhancer elements.

4.4 The genetic background determines OR expression levels independent of odour environment.

Genetic variation will, undoubtedly, have an effect on the expression of the OR genes. However, genetically divergent mouse strains also produce different odours in their urine[326, 327] and amniotic fluid[51]. Therefore each strain of mouse, when housed in homogeneous groups, is exposed to a unique pre- and post-natal olfactory environment. Odour exposure has been shown to alter the life-span of OSNs in an activity dependent manner[286, 288, 292]. Thus, the observed changes in OR gene expression between the different strains could be the result of their genetic makeup and/or the odour environment they are exposed to. I therefore designed an experiment to dissect the genetic from the environmental contribution to OR expression regulation.

B6 and 129 embryos were transferred to F1 mothers and allowed to develop in this equivalent *in utero* environment. After birth, B6 litters were cross-fostered to B6 mothers and 129 litters to 129 mothers. Further, B6 litters received a single 129 pup, and 129 litters received a single B6 pup. In this setting, each litter has a characteristic olfactory environment, but one animal (the *alien*) has a different genetic background (Figure 4.12). This arrangement was maintained after weaning into single-sex groups, such that each alien was caged with four animals of the alternative strain. At 10 weeks of age, the WOM was collected for six alien animals and six cage-mates and RNA-sequenced (Tables B.1 and B.2 in Appendix B).

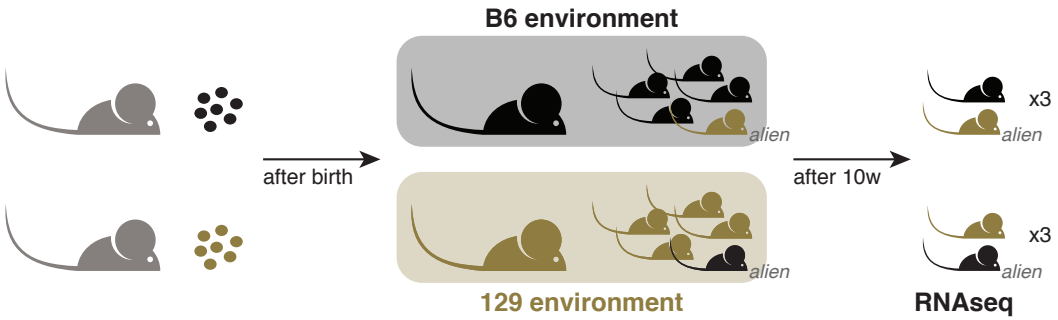


Figure 4.12 – Experimental design to dissect genetics from environment. B6 (black) or 129 (brown) embryos were transferred into F1 recipient mothers (grey). Once the pups were born, the B6 litters were cross-fostered to B6 dams and the 129 litters to 129 mothers. At the same time, a single B6 animal (the *alien*) was introduced to the 129 litter and *vice versa*. These groups were conserved after weaning, in 4:1 strain proportions. At 10 weeks of age, the WOM was collected from the alien animal along with a cage-mate. Three biological replicates were performed for each setup.

Evaluation of the OR gene expression repertoire revealed that animals clustered in two groups, clearly defined by the genetic background of the animals (Figure 4.13A). In other words, all the B6 animals displayed a similar OR gene profile, irrespective of their olfactory environment. Consistent with this, the correlation coefficients for any two B6 samples was on average 0.97, with no significant difference between the environments (p-value = 0.09, t-test); in contrast, the correlations for any B6 with a 129 sample had a mean of 0.89, which is significantly lower (p-value = 3.8e-12, t-test). 507 OR genes, among 5,475 genes were DE between these mice when grouped by strain, a similar number to the B6 vs 129 comparison when in their own olfactory environments (Figure 4.6). In striking contrast, across the whole transcriptome, only two genes showed differences in expression according to environment, both of which were OR genes (Figure 4.13B-C). For one of these ORs, the B6 animals in the 129 environment showed an expression pattern that resembles that of the 129 genome (Figure 4.13B); the other, however, had similar expression levels for both strains in their cognate environments,

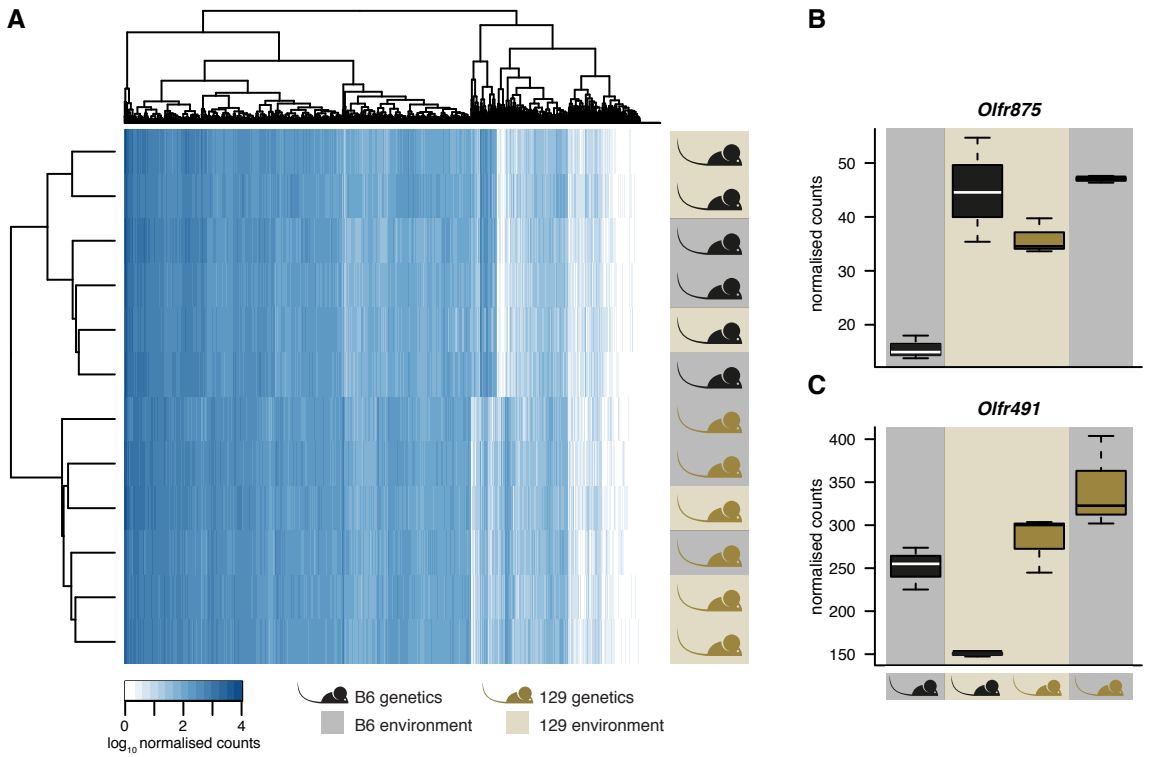


Figure 4.13 – OR expression is determined by the genetic background. A) Heatmap of the normalised OR expression for all the B6 (black) and 129 (brown) animals. To the right, the colour of the mouse indicates its genetic background and the colour surrounding it the type of environment it was exposed to. The samples cluster in two well defined groups, characterised by the strain of the animals. There is no clustering between the different environments. **B)** Two OR genes that showed altered expression levels upon differences on the olfactory environment. The colour of the box indicates the strain of the animal and the background shade the environment.

but when the B6 animals were switched to the 129 environment, the expression was downregulated (Figure 4.13C). No other genes reached statistical significance suggesting that environmental changes have very little influence on the regulation of gene expression in the main olfactory system. These data demonstrate that the WOM transcriptome is mostly influenced by direct genetic effects and the indirect effect of the olfactory environment is minimal and perhaps restricted to only a couple of OR genes. What's more, these results imply that the overall abundance of each OR type is independent of its activity or responsiveness to odorant cues.

To further test this, I sequenced the WOM of newborn B6 pups[1]. These animals were mostly restricted to their *in utero* environment and had only brief interaction with outside odorants. Therefore, the observed expression of the OR genes had minimal influence from olfactory stimulation. The expression of OR genes in these pups could

[1]Tissue collected by Darren Logan.

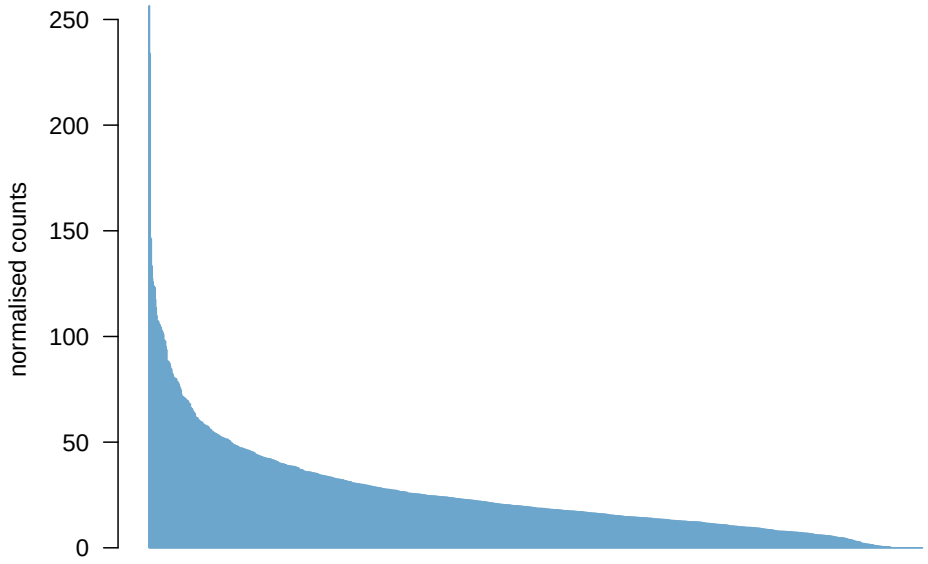


Figure 4.14 – OR expression in B6 pups. OR expression levels in B6 E19.5 newborn animals. Despite the abundance of OR expression being lower than in adults, there is a clear distribution of unequal proportions of the different OR genes, with some much more abundant than others.

be readily detected in the RNAseq data. From the whole OR repertoire, 1,198 (95.9%) genes showed evidence of expression. The olfactory system is still developing at birth and there is rapid growth and a steep increase in the number of OSNs during the first postnatal weeks[26]. Consistently, the expression levels of the receptor genes were low compared to adult animals. Nonetheless, a similar distribution of unequal proportions of different OR genes was readily observed (Figure 4.14). The median OR expression was 18.76 normalised counts, and only 24 receptor genes had abundances higher than 100 normalised counts. Thus, the differential proportions of OSNs expressing particular OR genes are present before any significant olfactory stimulation has occurred, suggesting that it is not dependent on the activity of the OSNs.

Two other observations support that the expression distribution observed is not influenced by the activity of the receptor genes. First, analysis of the expression of receptors that are pseudogenised and do not produce functional proteins revealed a similar distribution, of some genes expressed more abundantly than others (Figure 4.15). And second, the collection of OR genes that have identical coding sequences between different strains –and therefore identical proteins– very often occupy varying positions in the distribution. 36.3% of the OR genes that are identical between B6 and 129 are significantly DE; similarly 44.8% of the identical receptors between B6 and CAST are expressed at different levels.

Thus, the activity of the receptor protein itself has no influence on the final expression

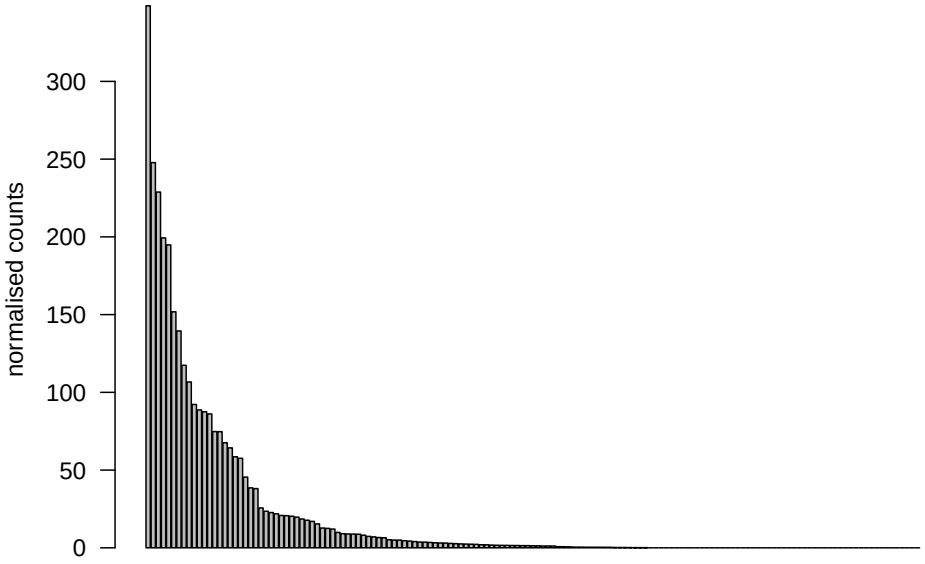


Figure 4.15 – Pseudogene OR expression in B6. OR expression levels for all those receptors annotated as pseudogenes in the B6 genome. Some genes are expressed at much higher levels than others, suggesting that the abundance of each receptor is independent of the activity of the protein.

level of the gene. All data indicates that the observed differences in OSN composition in the MOEs of different mouse strains are driven by changes in their genetic architecture; the state of the regulatory sequences present in each genome dictate the final proportion of OSNs that express each receptor type.

4.5 OR expression is controlled in *cis*.

Several enhancer elements have been identified that regulate the probability with which OR genes from nearby clusters are chosen[183, 186, 187]. For these, it has been demonstrated that their regulatory activity acts in *cis* and do not influence the expression of the homologous alleles on the other chromosome[183, 184, 186]. However, 3C experiments have indicated that there are interchromosomal interactions between different enhancer elements[187]. Further, I have now shown that the expression level of each receptor gene depends on the genetic context. In order to determine if the observed differences in expression are the product of *cis*-acting elements, I analysed available WOM data from B6×CAST F1 hybrids[2]. When compared to the parental strains, these hybrid animals provide information about the regulatory elements affecting gene expression. For all those genes that are differentially expressed between the parental strains, the corres-

[2]Raw sequencing data was kindly provided by Sophia Liang.

ponding expression of each allele can be determined in the F1 by interrogating variable positions. If the ratio observed between the parental expression levels is the same as that of the two alleles, it can be inferred that expression is regulated by elements acting in *cis*; on the other hand, if the expression of the two alleles in the F1 is no longer different, then regulation is occurring in *trans*[328].

For the OR repertoire, 1,018 (81.5%) of the genes have at least one SNP reported for CAST that is covered by sequencing fragments. At each of these positions, I calculated how many sequencing fragments pertained to each allele. Then, I used the ratio of B6 to CAST counts to deconvolve the total gene expression into allele-specific expression for each OR gene (see Appendix A for detailed methods). Figure 4.16A shows the fold-change in OR gene expression in the parental strains versus the corresponding fold-change of the alleles in the F1; most genes lie along the 1:1 diagonal, indicating that the expression levels observed in the animals with pure genetic background (B6 or CAST) are conserved at the allele-level in the F1 and, therefore, must be controlled in *cis*. In other words, the expression of the two alleles within the F1 animals is the summation of the patterns present in the parents (Figure 4.16B).

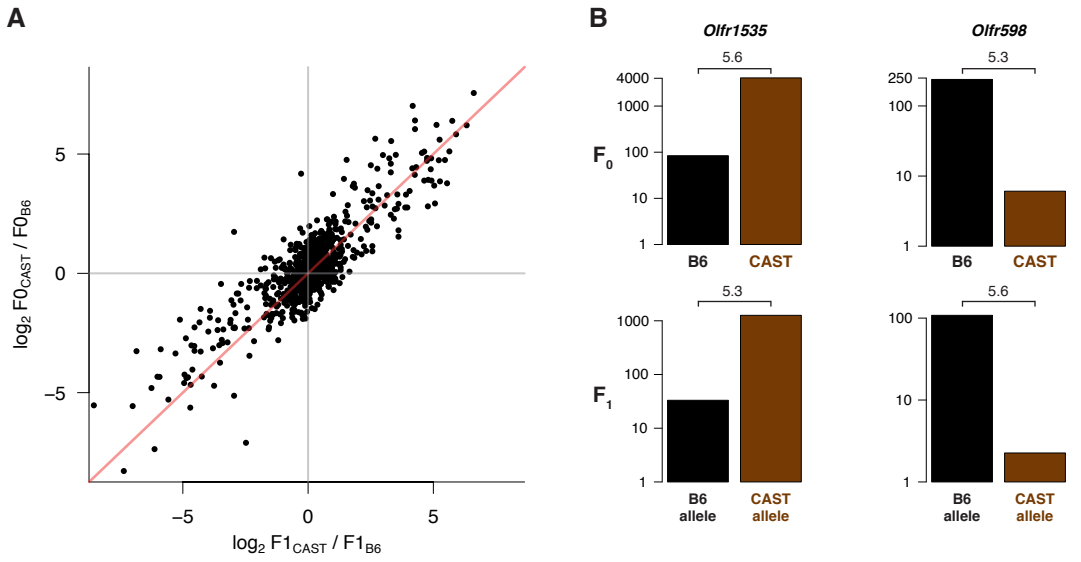


Figure 4.16 – OR expression is regulated in *cis*. A) Scatter plot of the fold-change in OR expression between the CAST and B6 parental strains (y-axis) versus the allelic expression for each strain in the F1 (x-axis). Most genes lie in the 1:1 diagonal (red), which indicates that the proportions observed in the parents are conserved in the F1. This occurs when expression is regulated by *cis*-acting elements. **B)** Examples of two OR genes that are differentially expressed in the parental strains (top row). The expression in the F1 hybrids was decomposed into the contribution from each allele, and this is plotted on the bottom row. The $\log_2$ fold-change is indicated in each case and these are equivalent for both the F_0 and F_1 data.

All together, these data provide a comprehensive landscape of the OR repertoire transcriptome from genomes with varying levels of divergence. A great proportion of

the receptor genes are susceptible to differences in their expression levels. These correlate with disparity in the number of OSNs that express each OR gene. Further, the changes are directly determined by the genetic architecture of the animal but not by their olfactory environment or the activity of the receptor protein. Based on the expression pattern of OR genes in an F1 hybrid, it can be inferred that the vast majority of regulation of OR gene choice is conducted by elements acting in *cis*. Genetic variation within these –poorly defined– regulatory elements correlates with differential OR gene expression. Together, these data are consistent with a model where genetic variation in regulatory elements alters the probability with which each receptor is chosen, thus resulting in the creation of a highly diverse mosaic composition of the MOE, with varying proportions of each OSN type.

Chapter 5

Olfactory stimulation alters the OR repertoire.

I have previously shown that the proportion of OSNs expressing any given OR is conserved between animals of the same genetic background, but differs considerably when genetic variation is introduced. Also, I have demonstrated that the observed changes are due to *cis*-acting regulatory elements and that the olfactory environment has very little effect on the WOM transcriptome. However, several studies have shown that neurones that are activated by their cognate ligands have increased life-spans[286, 288, 292]; with time, their longer survival rates translate into an enrichment in the neuronal population, compared to those OSN types that are mostly inactive[292]. Thus, one might expect that the OSNs that express receptors responsive to the odorants that are differentially produced by B6 or 129 animals, or by males and females, should be overrepresented in one strain or sex. But my results in Chapter 2 and 4 show only a few OR genes are differentially expressed in any of these comparisons. However, it is also well known that persistent exposure to any given odorant results in adaptation, where OSNs are desensitised and inhibit their responses to such olfactory stimuli[245]. Hence, could it be that the lack of DE OR genes between males and females, or between B6 animals living in a 129 olfactory environment, is the result of olfactory adaptation?

5.1 Acute but not chronic odour exposure affects OR expression levels in the WOM.

To investigate if the overall OR expression in the WOM is susceptible to change upon olfactory stimulation, I designed a set of odour exposure experiments. I chose four different odorants for which receptor-ligand interactions have been well characterised: *Olfr50* (*I-D3*) responds to (R)-carvone[224], *Olfr151* (*M71*) to acetophenone[216], *Olfr2* (*I7*) to heptanal[224] and *Olfr73* (*mOR-EG*) to eugenol[215] (Figure 5.1A). These four compounds were dissolved in mineral oil in equimolar proportions, for a final concentration of 1mM each. The odour mixture was used to stimulate B6 male and female mice, with two different presentation paradigms: 1) A *chronic* exposure, that consisted of adding the mixture to a cotton ball placed inside a tea strainer, and left in the animal's cages 24 hours a day; the mixture was replaced fresh every day. 2) Or an *acute* exposure design, where the mixture was added to the drinking water supplied to the animals (Figure 5.1B). In the chronic paradigm, the odorants were present in the environment uninterruptedly, while in the acute set-up, the animals could smell the mixture only when they approached the bottle to drink. For both experiments mineral oil was used as a control.

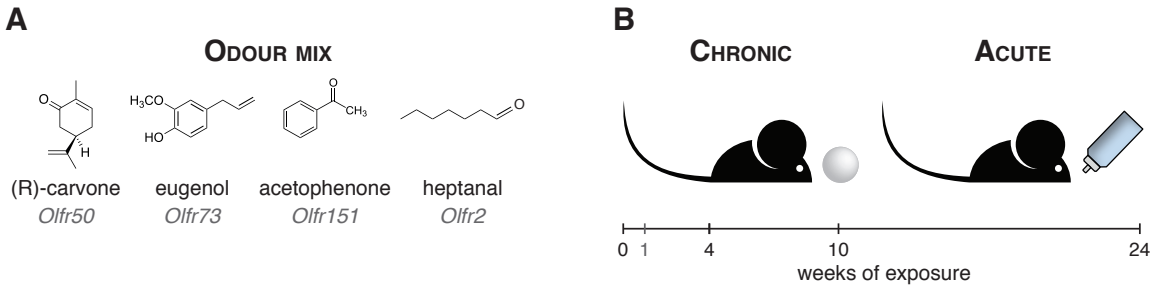


Figure 5.1 – Odour exposure experimental set-up. A) Four different odorants were used to stimulate B6 mice, as an equimolar mixture: (R)-carvone, eugenol, acetophenone and heptanal. The chemical structure of each is shown, along with one OR they activate. **B)** The odour mix was presented chronically, by applying it to a cotton ball that was left inside the mice cage 24-hrs a day; or acutely, by adding the odour mixture into the drinking water. In the acute paradigm, the animals smell the odorants only when they approach the bottle to drink. For both experiments, WOM was collected after 4, 10 and 24 weeks from the start of the treatment. For the acutely exposed animals, a further 1-week time point was included.

The odour-exposure was started from birth. I then dissected the WOM at different time-points, and assessed the expression of the OR genes expected to respond to the individual odorants, by TaqMan qRT-PCR. As *Olfr151* is a pseudogene in the B6 genome it was not included in the analysis. When the odour mixture was presented in an uninterrupted manner, no changes could be detected in the overall WOM expression of any of the three receptor genes in the odour-exposed animals compared to controls (Figure 5.2). However, the animals that were exposed acutely to the odorants, showed a

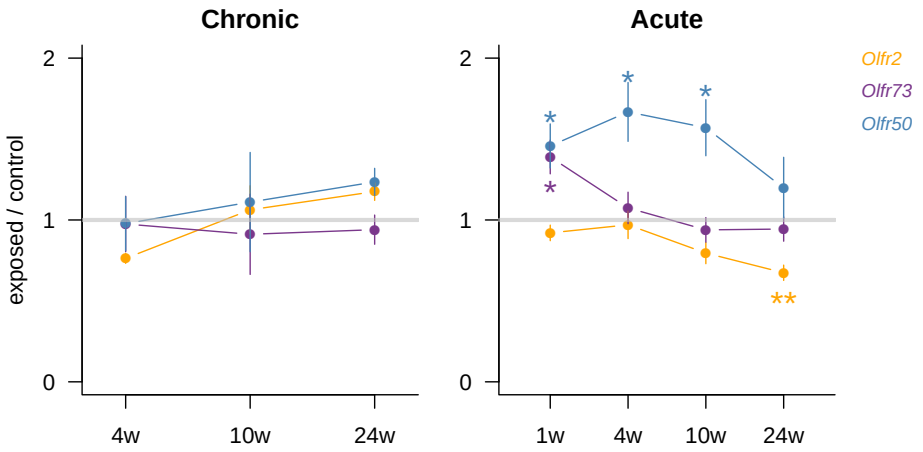


Figure 5.2 – OR expression is altered with acute stimulation. Expression estimates obtained with TaqMan qRT-PCR assays for *Olfr2* (orange), *Olfr73* (purple) and *Olfr50* (blue) at different time-points (in weeks). The ratio between the odour-exposed and the control animals is plotted. For the group exposed chronically (left) to the odour mixture, no changes were detected for any of the genes at any point. In contrast, the animals exposed acutely (right) showed significant changes for several genes. Error bars are the SEM. * P<0.05 ** P<0.01 (y-yesy, FDR < 5%). Chronic: n = 3-5 animals per group, for 4 and 10 week time-points; 9-10 for 24 weeks. Acute: n = 8-13 animals per group, per time-point.

consistent and significant upregulation of *Olfr50* at all time-points except at 24 weeks. Also, *Olfr2* was significantly downregulated at the latest time-point tested, and *Olfr73* was upregulated very early but this change was not maintained in later stages (Figure 5.2). Thus, these data suggest that OR gene expression is susceptible to dynamic change by exposure to specific odorants, but only when the stimulation is intermittent. However, it is also possible that the stimulation achieved by delivering the odorants in the water is substantially different to the chronic exposure paradigm, not only in frequency but also in intensity.

To assess the proportion of the OR repertoire that was affected by the acute exposure treatment, I sequenced RNA from six control and six experimental samples from the 24-week cohort (Tables B.1 and B.2 in Appendix B). Differential expression analysis revealed 36 OR genes were significantly regulated (FDR < 5%) upon exposure to the odour mix, with similar numbers of receptors more or less abundant in the treated animals (Figure 5.3). Most of the DE genes had small fold-changes; only one third showed differences in expression of 1.5 fold or more.

To further validate these changes, I selected seven genes with the biggest differences for which specific TaqMan probes were available, and tested their expression in all the sequenced samples plus 3 extra controls and 7 additional exposed animals. All the tested genes were statistically significant (t-test, FDR < 5%) and the direction of the changes was concordant with the RNAseq data (Figure 5.4). I then tested these genes in the

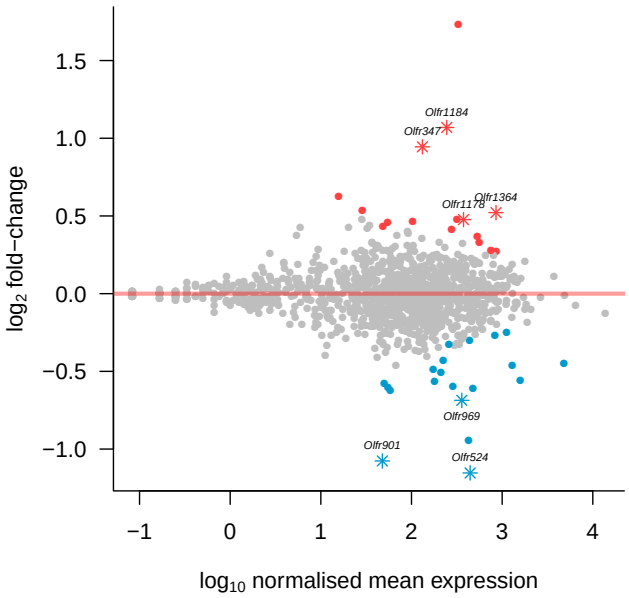


Figure 5.3 – DE OR genes in acutely exposed mice. Mean normalised expression estimates for the OR repertoire are plotted aginst their fold-change in the animals exposed acutely to the odour mix (for 24 weeks) versus controls. In all, 36 receptors are significantly differentially expressed (FDR $< 5\%$), with 16 being more (red) and 20 less abundant (blue) in the treated animals. The red line indicates equal expression in both groups. Several DE genes were selected for further validation; these are plotted as stars and their gene name is indicated. $n = 6$ animals per group.

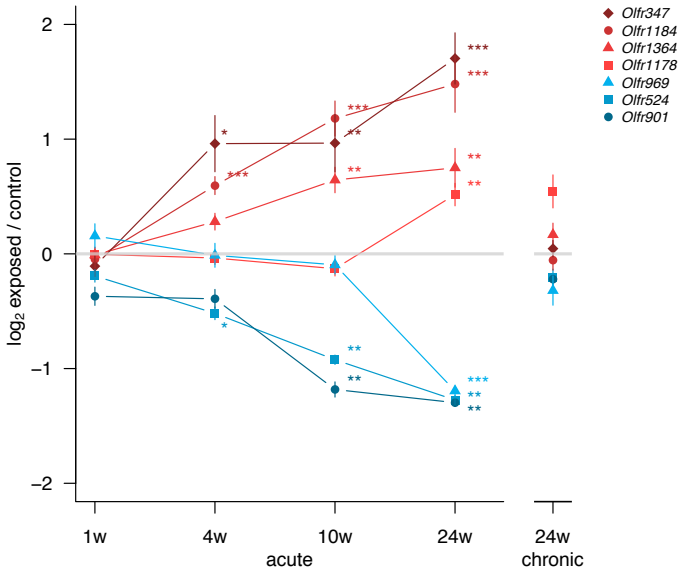


Figure 5.4 – ORs regulated by odour stimulation change in a time-dependent manner. Fold-change of the expression of seven DE OR genes in odour-enriched animals versus controls, assessed by TaqMan qRT-PCR. The ratio between groups is presented for each gene at different time-points after the start of the odour exposure, for the acutely treated mice. All seven genes are significant after 24 weeks of treatment, but some are different as early as 4 weeks. To the right is also the fold change for the same genes in the animals treated chronically during 24 weeks. No significant changes were detected. Error bars are the SEM. *P < 0.05 **P < 0.01 ***P < 0.001 (t-test, FDR $< 5\%$). $n =$ 8-13 animals per group, per time-point.

samples from previous time points. After 4 weeks of exposure, three of the genes were already statistically significantly different from controls, and at 10 weeks five out of the seven receptors were clearly DE (t-test, FDR $< 5\%$). After only one week of treatment none of the genes showed significant changes, which was expected since the pups do not drink the odourised water at this stage. For five of the DE receptors, the difference in expression relative to controls increased with time, while the other two were unchanged until the 24 week stage; these might have slower change dynamics or might require a bigger change to be detectable by measuring expression levels in the WOM (Figure 5.4). Next, I tested these same DE receptors in the samples from the mice exposed chronically to the odour mixture for 24 weeks. None of the genes were significantly different in the exposed animals compared to controls (t-test, FDR $< 5\%$; Figure 5.4). Thus, I have shown that intermittent exposure to a set of odorants results in changes in the expression of several OR genes, that become accentuated with time. However, this is not observed if the odorants are constantly present in the environment, at least for these particular set of genes.

Finally, to assess the plasticity of the observed changes, I acutely stimulated a group of animals for four weeks, and then left them to recover for an additional six weeks (Figure 5.5). I then collected the WOM and tested the expression of the same seven OR genes by qRT-PCR. As seen before, after 4 weeks of exposure, three of the seven DE ORs were significantly different and these changes increased and became more significant at 10 weeks. However, in the animals that were returned to pure water for the last 6 weeks, the expression of all the receptors was not different from controls (Figure 5.5). Thus, this data indicate that the changes in expression of certain OR genes upon odour exposure are plastic and require constant stimulation to be maintained.

5.2 Differential regulation of OR genes is odour-specific.

To better understand the effects of individual odorants on the expression of the OR repertoire, I repeated the acute odour exposure experiments but supplementing the water with (R)-carvone or heptanal alone, or with the combination of both. Controls were kept with pure water. I collected the WOM after 10 weeks of exposure and tested the expression of the seven DE ORs identified previously, by TaqMan qRT-PCR. None of the genes were significantly DE in the animals exposed to (R)-carvone alone; a marginal

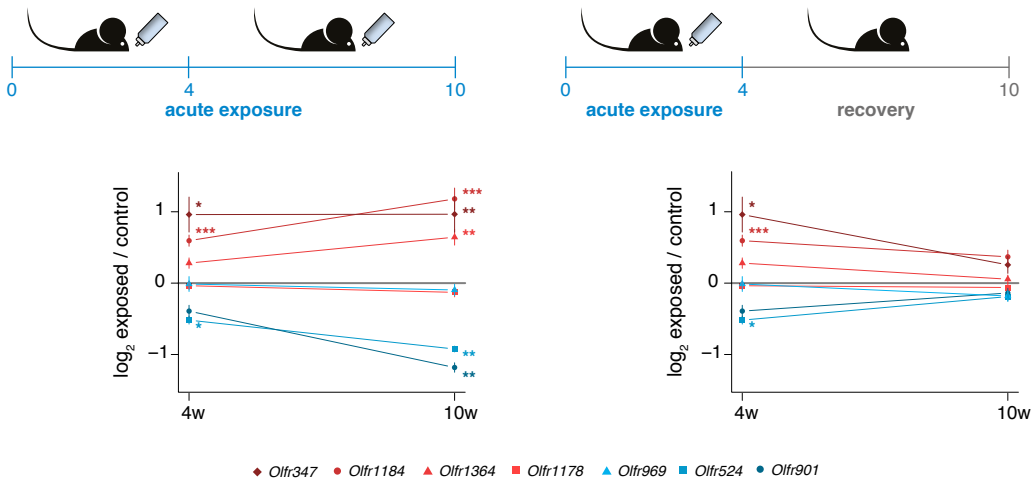


Figure 5.5 – Changes in OR abundance are plastic. Mice were acutely exposed to the odour mixture for four weeks. Then, they were returned to pure water for 6 weeks. The expression levels of the DE OR genes were tested at 10 weeks, by TaqMan qRT-PCR. The fold change between the treated and control animals are plotted for each gene. On the left is the same data from Figure 5.4; after 4 weeks of treatment three of the genes are significantly different from the controls and at 10 weeks these changes become more pronounced. On the right, the expression values returned to control levels in the animals left to recover, with no significant differences for any of them. Error bars are the SEM. *P $<$ 0.05 **P $<$ 0.01 ***P $<$ 0.001 (t-test, FDR $<$ 5%). n = 8-9 animals per group.

upregulation was observed for *Olfr50*, the carvone cognate receptor, but the difference was not significant after correcting for multiple testing. However, four of the seven tested receptor genes were significantly different in the animals exposed to heptanal, or to the combination of both odorants (t-test, FDR $<$ 5%; Figure 5.6).

To fully characterise the changes occurring in each of these experimental groups I performed RNAseq (Tables B.1 and B.2 in Appendix B). Differential expression analysis

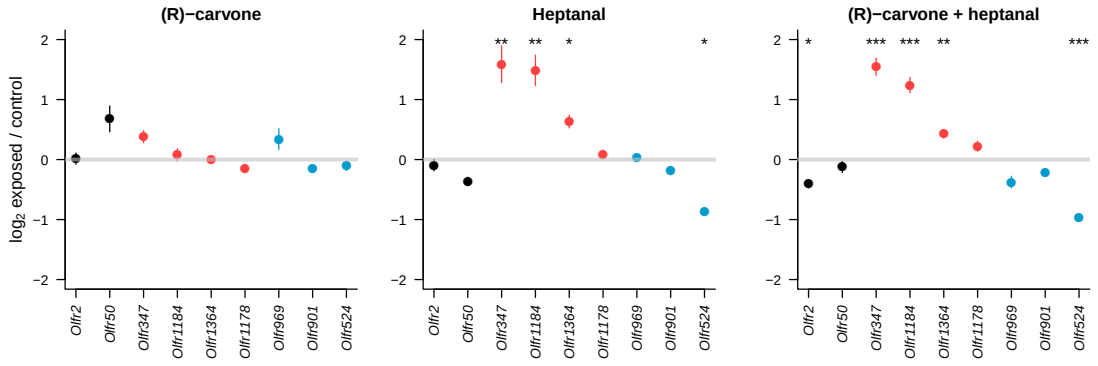


Figure 5.6 – OR genes respond to odour stimulation in a specific manner. The fold-change of the expression values for 7 OR genes previously shown to be DE upon odour-exposure are plotted for animals treated with only (R)-carvone, only heptanal or the combination of both, as assessed by TaqMan qRT-PCR. The cognate receptor for the two odorants are in black, DE genes up- and down-regulated in treated animals are in red and blue respectively. (R)-carvone has little effect on these genes and the changes are not significant. Heptanal, however, affects four of the seven genes. Error bars are the SEM. *P $<$ 0.05 **P $<$ 0.01 ***P $<$ 0.001 (t-test, FDR $<$ 5%). n = 6 animals per group.

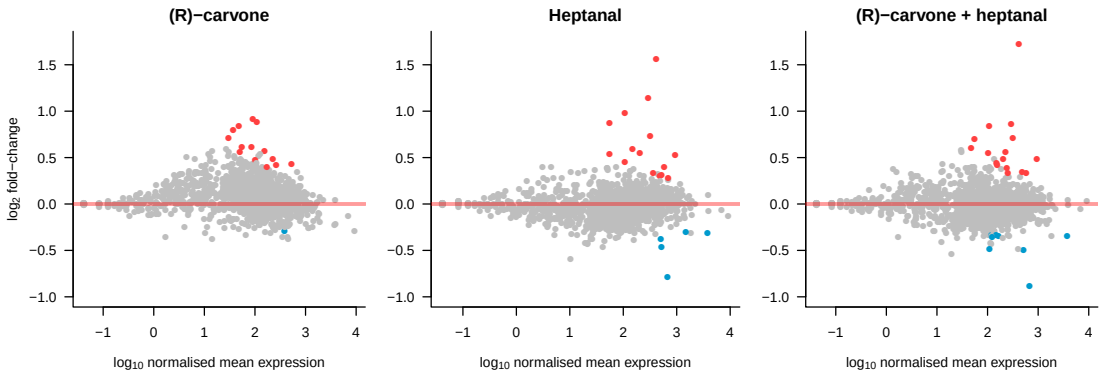


Figure 5.7 – DE OR genes in mice stimulated with different odorants. Plots of the normalised mean counts for all the OR genes versus their fold change in the experimental groups compared to controls. Each of the groups exposed to (R)-carvone, heptanal or the combination of both are presented. Statistically significant DE genes (FDR < 5%) are highlighted in red if higher in the experimental group, and in blue if lower. The red line indicates expression is not different to controls. $n = 6$ animals per group.

revealed that, in all, 43 ORs were significantly DE in at least one of the conditions (FDR < 5%), and the majority of these (74.4%) were upregulated in the odour-stimulated animals. Exposure to (R)-carvone or heptanal resulted in the change in expression of 15 and 20 OR genes, respectively (Figure 5.7). These sets of receptors were almost completely independent, with only one significantly regulated in both groups (*Olfr538*; Figure 5.8A-B). The animals that were exposed to both odorants simultaneously showed significant changes for 24 OR genes, 15 of which were shared with the individually exposed groups (Figure 5.7). Interestingly, the great majority of these overlapped with DE genes in the heptanal group, and only 2 were shared with the (R)-carvone group (Figure 5.8A). Thus, the data suggests that exposure to (R)-carvone and heptanal alters the global expression of some OR genes in the WOM, and these changes are odorant dependent. But, when both odorants are presented in combination, the effects exerted by heptanal overpower those of (R)-carvone, resulting in an expression profile that resembles more that of the heptanal group (Figure 5.8). Finally, I compared the DE ORs from both experiments, either using specific odorants or the mixture of four. Almost 40% of the ORs that showed significant changes when exposed to all four odorants were also altered in one or more of the groups exposed to (R)-carvone, heptanal or their combination.

All together, I have identified sets of OR genes that are regulated by exposure to odorants, when the stimulation is intermittent but not when the odour cues are present permanently. The receptors that are altered are dependent on the odorant used and the changes are reversible. Thus, these data are consistent with the hypothesis that constant but interleaved activation of OSNs expressing particular ORs results in a differential proportion of such OR genes in the overall WOM transcriptome. Once the stimulation

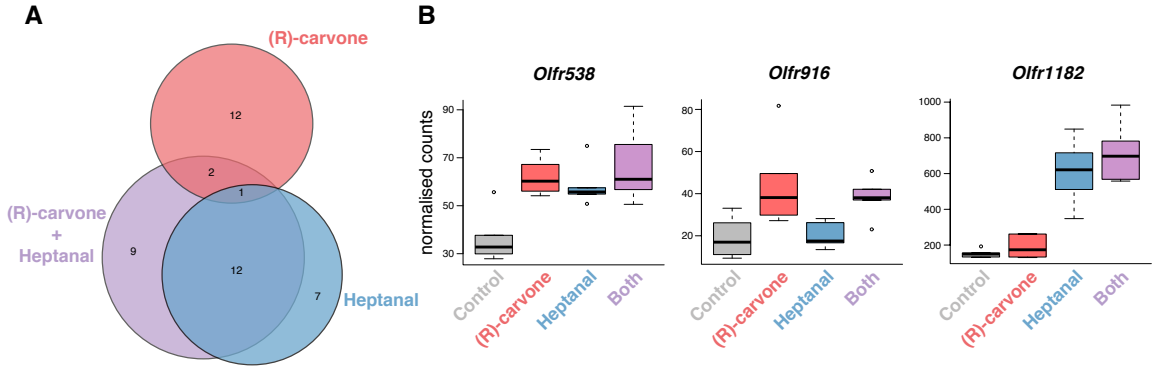


Figure 5.8 – Different ORs respond to specific odorants. A) Venn diagram showing the proportion of DE genes in each of the experimental groups exposed to (R)-carvone, heptanal or their combination (both). Genes regulated by exposure to (R)-carvone are different than those regulated by heptanal. **B)** Examples of a significant DE OR gene that is influenced by both odorants (*Olfr538*), by (R)-carvone but not by heptanal (*Olfr916*) and *vice versa* (*Olfr1182*).

ceases, the changes are reversed and the expression of the OR repertoire returns to the stable state dictated by the genetic composition of the animal.

Chapter 6

Discussion and future perspectives

In this dissertation I have explored the dynamics of OSN diversity, as measured by olfactory receptor expression, and the effects of genetic and environmental factors on its regulation. For this, I utilised an RNAseq-based approach that allowed me to study the complete OR repertoire. I have shown that the transcriptional profiles obtained via RNAseq from whole tissue extracts are accurate and highly reproducible, and outperform other technologies available. Further, the high-throughput and unbiased character of the technique allowed the generation of a comprehensive catalogue of the transcripts present in the olfactory system, both known and novel; and the generation of full-length gene models for hundreds of OR and VR genes. The combination of RNAseq with FACS and single-cell technologies resulted in a precise characterisation of the molecular profile of the OSN transcriptome. Moreover, it allowed the discovery of novel subdivisions of mature OSNs. Importantly, the study of single OSNs permitted me to assess the wide-held belief that OR expression is monogenic and monoallelic, hereby directly proven.

From the data, I can conclude several things. First, expression levels of OR genes in WOM samples are an accurate reflection of the number of OSNs in the MOE that express particular receptors. Thus, the transcriptional profiles inform on the proportions of the different OSN types found in the neuroepithelium. Second, such diversity of OSN types is stereotypical in animals of the same genetic background, irrespective of sex and (largely) of age. Third, the presence of genetic variation results in high divergence of the relative proportions of different OSN types, with most being susceptible to altered abundance based on their genomic context. Fourth, the final distribution of OSN diversity is controlled by genetic elements that act in *cis*, and is not affected by sustained alterations of the olfactory environment. And fifth, the persistent but interleaved presentation of olfactory stimuli alters the abundance of a subset of OSN types,

in a time-dependent, odour-specific and reversible manner (Figure 6.1).

6.1 Understanding the mouse olfactory system by RNAseq.

High-throughput RNA-sequencing has greatly advanced and developed in the last few years. With the sequencing costs dropping, RNAseq has substituted the use of microarrays and has become routine for transcriptional profiling. I have exploited the strengths of RNAseq to characterise and better understand the transcriptional dynamics of the mouse olfactory system. In this dissertation, I have demonstrated the accuracy and reproducibility of the expression estimates obtained with this methodology, and its superiority when compared to other techniques such as a greater dynamic range of expression values and better correlation with qRT-PCR expression estimates.

During the course of my PhD, several groups published results from similar experiments to mine[329–331]. Shiao et al. performed RNAseq in WOM samples of male and female BALB/c mice and concluded that males have overall higher expression of OR genes[329]. But they failed to notice that males also have higher expression of all the canonical markers of mature OSNs and, therefore, it is likely that the observed differentials are only a product of varying proportions of OSNs in their whole tissue samples. Additionally, they sequenced only one sample of each sex (that was the pool of three individuals), which makes it very difficult to test for differential expression with confidence, since any observed differences could be the result of technical variation with no biological relevance. Indeed, both my data and that of Kanageswaran et al. [330] failed to identify any convincing differential OR expression by sex.

A similar situation occurs in the experimental design of Kanageswaran et al.[330]; they sequenced several biological replicates of MOE samples from CD1 and B6 mice, but only a single replicate of FAC-sorted OSNs from OMP-GFP animals. Thus, the comparisons of the transcriptomes of the OSNs versus the whole tissue are underpowered and are also flawed since the genetic background was not controlled. Finally, Shum et al. presented WOM RNAseq data of two adult B6 females and reconstructed OR gene models using a strategy similar to mine; however, their sequencing depth was much lower. An analysis of their reconstructed gene models led to the proposal that the presence of introns leads to higher expression levels[331]. However, the authors did not recognise that the genes they classified as *intronless* were so because they did not have enough

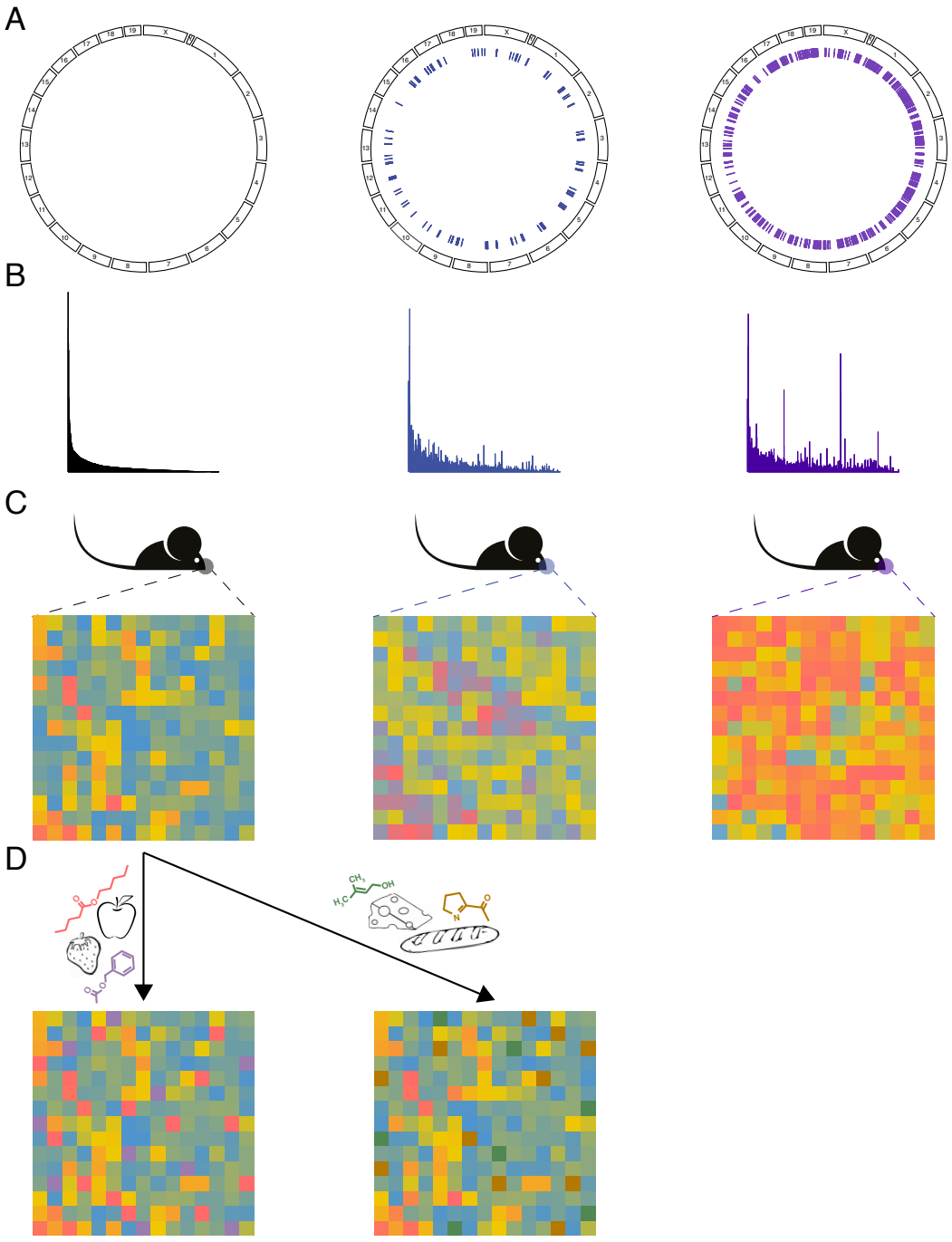


Figure 6.1 – **Genetic and environmental regulation of OSN diversity.** Model on the impact of genetic and environmental factors on OR expression and OSN type abundance. **A)** Representation of the mouse genome, where all chromosomes are arranged in a circle. The chromosome number is indicated. As an inner circle is a depiction of genetic variation events, each as a vertical line. **B)** The first genome (left) produces a particular distribution of expression for all the OR genes, ordered by decreasing abundance order. For the second (centre) and third (right) genomes, OR genes are ordered the same as in the first. The presence of genetic variation results in OR expression profiles that are different from each other. Greater amounts of variation (right) result in a more divergent profile. **C)** The unequal expression levels for different OR genes results in unequal numbers of OSNs expressing such receptors. Each square represents a different OSN and the colour indicates the particular OR gene it expresses. Thus, each mouse with a unique genome has a unique pattern of OSN diversity in its nose. **D)** The proportion of each OSN type is amenable to modification upon olfactory stimulation. On the left is represented a country mouse that feeds on fruit and seeds; constant exposure to odorants from fruits results in the enrichment of OSNs that express ORs that recognise such odorants (pink and purple). On the right is a city mouse that instead feeds on cheese and bread and therefore has more OSNs that express ORs activated by the molecules in these foods (green and brown).

depth to reconstruct their full gene models. Indeed, from all the OR genes proposed to be intronless, 91.9% have introns in the reconstructed models I presented. Thus, the low expression levels were the cause of the lack of introns in their models, and not the other way around.

An important advantage of RNAseq is that it does not depend on the genome annotation; this has allowed me to discover novel genes that, to date, remain unannotated in the mouse genome. Several hundred loci showed evidence of expression of multi-exonic structures, many of which contained protein features and domains. The surprisingly high number of such novel genes indicates that despite the high quality of the sequence and annotation of the mouse genome, the use of specialised tissues that are often not considered in gene annotation pipelines are a source of additional information. It is likely that many of the putative genes I have defined are specifically expressed in the olfactory system and serve specialised functions; examples are the two lipocalin genes (*Lcn16* adn *Lcn17*) that were validated as true genes, and are odorant binding proteins. These two novel genes can be found in the same orientation in the rat genome, but synteny is disrupted in the primate lineage and there are no orthologues present in primates or humans.

RNAseq combined with other experimental strategies, such as cell sorting based on expression of particular marker genes, provide a powerful strategy to deconstruct complex tissues such as the MOE. In collaboration, I utilised this approach to characterise the transcriptional profile of the OSNs only, excluding the other cell types present in WOM samples. By differential expression analysis, I defined a large list of genes that are specific to the neurones and provide information on their molecular processes and pathways. These will be useful as a reference for future studies, since discriminating neuronal processes from supporting functions is very informative on the nature of the phenomena being studied.

Perhaps more interesting was the finding that the expression of *Omp*, the canonical marker that defines mature OSNs, is not expressed in a continuum but, instead, segregates into two discrete populations. The analysis of the genes that differentiate these two subpopulations revealed that while both are mature OSNs, the GFP^{low} cells are slightly less mature than the GFP^{high} cells. A recent study characterised the temporal expression of OR genes, *Adcy3* and *Omp* in differentiating precursor cells into mature OSNs. Indeed, *Omp* was found to be the last gene to be activated, after both OR genes and *Adcy3* had been turned on[312]. Therefore, it is conceivable that the difference between OSNs expressing either low or high levels of *Omp* coincides with a discrete functional

event in the final maturation of the neurones. For example, it could be that the cells from the GFP^{high} population have successfully established the negative feedback that ensures OR singular expression; the release of the unfolded protein response might be necessary to achieve high *Omp* expression. Alternatively, it could be that the distinction reflects the successful innervation of a glomerulus in the MOB. In either case, this finding likely marks an event in the functional maturation of OSNs and deserves further study.

Probably the most important advantage of RNAseq as a technology, is that it provides a comprehensive and unbiased profile of all the genes expressed in a particular sample. As such, it is a very attractive strategy to study the basic questions of OR gene expression regulation. Ever since the discovery of the OR genes and the study of their expression in the MOE, it has been assumed that they are expressed in a monogenic and monoallelic fashion in every OSN[66, 67, 69, 72]. However, all the evidence supporting monogenic expression of ORs stems from testing coexpression of a few combinations of two receptors but the full OR repertoire has never been tested[70, 182]. Therefore, the study of the transcriptome of single OSNs provides an unbiased method to account for *all* the different OR genes that are transcribed in a particular neurone. By sequencing 21 different individual OSNs, I was able to identify abundant OR expression of one OR gene in 19 of these. Additional receptor genes showed evidence of expression, but at very low levels. Indeed, taking together all the sequencing data supporting OR gene expression, in each OSN over 98.1% was concentrated on a single OR, with the remaining scattered across a few other receptors. It is not clear whether these low-abundance OR genes are biologically meaningful or whether they represent leaky transcription that has no impact on the sensing capabilities of the OSN. It has been shown that genes that are expressed at very low levels do not correlate with protein expression[305]; therefore, it is possible that only the abundant OR translates into protein. Additionally, low expression levels of some OR genes could be observed in other various cell types, which suggests that their expression is not related to olfactory function. However, until proteomic techniques match the sensitivity of the transcriptomic methodologies, this will remain unresolved. To date, studies of the membrane proteome of the cilia of OSNs have been able to identify only a few dozen OR genes, due to their low expression in WOM preparations[332].

Based on these data, the 'one neurone - one receptor' rule of OR expression is supported, for the first time, on a scale that accounts for each and every receptor annotated in the mouse genome. However, this is based only on 19 OSNs, which is far from representative sampling of the diversity of neurones present in the MOE. Thus, I cannot

rule out that some OSNs indeed express several receptors at high levels, but were not captured in the limited sample analysed here. Furthermore, the stringent QC criteria applied selected against the inclusion of neurones that contained two or more OR genes expressed at high levels, because some were clear cases of carry-over from other samples; therefore, it was not possible to ascertain the validity of the observed expression. In order to explore further the possibility that some OSNs express two ORs, collaborations are in place to obtain OSNs from mouse lines that express a specific receptor gene along a reporter fluorescent protein. By sequencing the population of neurones that express a particular OR gene, we will be able to gain insight into the levels of expression of other receptors. If the data indicates that coexpression of receptors is likely, a single-cell approach could then be used to definitely prove that this occurs within a single OSN.

The OMP-GFP animals used for the single-cell RNAseq experiments are in a mixed B6×129P2 genetic background. Using the SNPs present in the exons of the abundant OR genes, it was possible to infer the allele expressed in each OSN. By examining the sequencing data directly, I was able to confirm that OR expression is monoallelic, and that this is extremely tightly regulated. Over 99.7% of all the sequencing reads covering variable positions supported the expression of one of the alleles; the remaining reads presented any of the other three nucleotides, suggesting that these contain sequencing errors and low quality base calls, rather than being the transcription of the other allele. Therefore, the paradigm of expression of only one allele of the chosen OR gene holds true, at least based on the data from the small subset of OSNs tested.

Lastly, the combination of powerful technologies such as single-cell capture and preparation, along with RNAseq, allow for the discovery of minor populations of cells that are undetectable when bulk RNAseq is performed. In this way I was able to identify two OSNs that do not express any OR genes at high levels. Instead, they have very abundant expression of *Gucy1b2*, a soluble guanylyl cyclase, and *Trpc2*, the cation channel that is fundamental in signal transduction in VSNs. During the analysis of these data, a paper reported the existence of two different subpopulations of OSNs in the MOE that are positive for *Trpc2*[131]. Further characterisation of one of these –the type B cells– resulted in the identification of expression of *Gucy1b2* and suggested the lack of expression of chemoreceptors[322]. By profiling the complete transcriptome of two of these cells, I was able to identify over 50 genes that are not expressed in the canonical –OR-expressing– OSNs and, therefore, constitute the molecular fingerprint of this novel neurone type. Additionally, I confirmed that no known chemoreceptor is expressed at abundant levels. The *Trpc2*$^+$ OSNs have recently been shown to innervate glomeruli in

the MOB, which suggests that they might indeed be chemosensory cells relying olfactory information[131, 322]. Many of the components involved in the signalling pathway used to generate action potentials are present in these cells, and *Gucy1b2* is expressed at levels similar to OR genes; thus, it is possible that this guanylyl cyclase might be taking on the role of the receptor protein, but further studies are necessary to test this.

6.2 Almost all OR genes are expressed in the MOE.

The MOE contains a variety of OSN subpopulations, each defined by the particular receptor they express. The great majority express OR genes, but minor subpopulations also express TAARs or GC-D, and possibly several other subdivisions are yet to be characterised. In this dissertation I have presented, for the first time, the complete expression profile of the receptor repertoire in mice, with particular focus on the OR genes. Since the majority of the OR genes have been defined by computational methods, they lack evidence supporting their role in chemo-signalling. The most basic requirement for an OR to be implicated in olfaction, is that it is expressed in a sensory tissue such as the MOE. Therefore, the evidence that nearly the complete repertoire of putatively functional OR genes are indeed expressed in the WOM samples supports their involvement in transducing olfactory information. This is particularly relevant since extra-olfactory functions have been reported for some OR genes, that are not only expressed elsewhere[333], but have actually been shown to act in processes such as sperm chemotaxis[297, 298], muscle cell regeneration and migration[334] and serotonin release in gut cells[335].

Between 10 and 20 OR genes that are annotated as functional receptors lack expression data in any of the WOM samples. However, taking together the expression profiles of the different strains, only five lack expression in all the samples profiled. Thus, some of the receptors that are not expressed in a particular strain, might represent pseudogenes as a result of functional variation, or might be expressed at such low values that they were not detected. From the five OR genes with no unique counts in any sample, two are identical copies of each other (*Olfr247*), located ~8 kb apart; for these, all sequencing reads are multimapped and therefore it is impossible to know if one or the two genes are expressed. From the other three (*Olfr891*, *Olfr952* and *Olfr1061*), *Olfr952* has some multireads mapped and therefore could be expressed, while the other two do not. These could be cryptic pseudogenes, might be expressed at a different age or be present in an extremely low number of OSNs.

An important contribution of my work has been the construction of full-length gene models for a large number of OR genes. This has greatly increased the amount of information recovered for each receptor and, also, the amount of sequence that is unique. This has allowed me to estimate expression levels with more accuracy, since OR genes with close paralogs are systematically underestimated unless the more divergent non-coding regions of the genes are considered. But the additional sequence will also be very helpful for studies that are based on methodologies that rely on hybridisation approaches, such as qRT-PCR, NanoString nCounter, arrays and *in situ* hybridisation, one of the most popular for the study of OR expression. Genes that before were inaccessible are now available for study with probes specific enough to differentiate them from other receptor genes.

6.3 The MOE is a mosaic of OSN types.

The regulation of OR expression in OSNs is only partially understood. Achieving monogenic expression relies on a basal state of generalised repression of the whole OR gene repertoire; as the OSN reaches maturity, a single OR allele is activated[203, 205]. This is inefficient enough so that typically only one event can occur before a negative feedback mechanism ensures the process is shut down[209, 211]. However, how a particular receptor is chosen is still an open question. Very often, it is described as a *random* or *stochastic* process[202, 210], implying that any OSN can chose any of the 1,250 OR genes, and that any receptor has the same probability of being chosen. Contrary to this, each OSN has a restricted subset of the repertoire available for expression, depending on its location on the epithelium. Particular OR genes are expressed in restricted zones of the MOE[18, 66, 67, 78] and, therefore, only the OSNs located within those regions can choose them. Furthermore, it has been shown that different ORs are expressed in varying numbers of OSNs, with some being much more abundant than others[62]. This is at least partly influenced by the activity of enhancer elements and the number and organisation of transcription factor binding sites in the receptors' promoters[186, 188]. Thus, *random choice* is an unfortunate choice of words.

By profiling the entirety of the OR gene repertoire I have demonstrated that the expression levels of different receptor genes are highly variable, spanning at least four orders of magnitude. Importantly, I have also shown that the RNAseq expression estimates correlate with the number of OSNs expressing a particular OR gene; therefore, a highly expressed OR implies a high number of OSNs in the MOE expressing such a re-

ceptor. With this in mind, the unequal expression levels observed from the RNAseq data represent unequal proportions of each OSN type. This disparate distribution could be the result of two mechanisms, that might act alone or in combination: 1) differences in the frequency with which a particular OR gene is chosen, or 2) variation in the life-span of each OSN type such that, with time, those that live longer become more abundant in the overall population. It is unlikely that the second mechanism acts alone, given that unequal expression of ORs are already observed in newborn animals; in these, presumably not enough time has passed to allow survival dynamics to impact the proportion of each OSN type.

Based on the sequencing of several biological replicates from both male and female mice, I have determined that each individual expression pattern for the OR repertoire is exactly the same, as long as the genome remains unchanged (there are no OR genes in the Y chromosome). The rank correlation between different individuals is almost perfect (median rho = 0.98, p-value < 2.2e-16) which indicates that each receptor has equivalent values in the distribution, and the proportions of the different OSN types are preserved. Therefore, the contribution of each OSN type to the MOE's neural population is *determined* by the genetic architecture of the animal. The fact that males and females are indistinguishable, indicates that the OSN repertoire in the MOE is not influenced by the physiology or hormonal balance of the organism; nor is it altered by the differences in the olfactory environment produced by each sex. Furthermore, it is virtually unchanged at different ages within the controlled lab environment. The OR expression profiles of B6 animals of 10 or 24 weeks of age are equivalent and as highly correlated as between animals of the same age. This is consistent with a study of ageing female B6 mice, from 2 to 31 months of age, where the expression of 531 OR genes was assessed to find that only 4.3% of these were significantly differentially expressed[304]. But also, OSN abundance is unaffected by social and behavioural differences; for example, both sexes establish social hierarchies when group-housed[336], but no differences were evident between the different cage-mates. Thus, even though social interaction and behaviour are highly driven by olfactory cues, regulation of these processes is not achieved by differences in the receptor repertoire expressed in each animal. Instead, recent data suggests that internal state does alter olfactory perception, but by post-transcriptional mechanisms. Rather of changing the expression of the receptors, their activation is blocked by the influence of cycling hormones[337].

However, the distribution of OSN types observed in different strains of mice is highly variable (Figure 6.1B). Whereas B6, 129, CAST or OMP-GFP animals are all highly

correlated with those of the same strain, the differences between genetic backgrounds are large. More than 65% of the ORs were significantly DE between at least a pair of strains. This indicates that the great concordance observed between biological replicates is not the result of intrinsic stability or tight regulation of the expression levels of the OR genes. The different strains analysed in this dissertation are all inbred laboratory mouse strains. All the classical laboratory strains were derived from a small pool of founders from the *Mus musculus* (*M. m.*) *domesticus*, *M. m. musculus* and *M. m. castaneus* subspecies[321, 338]. Therefore, their genomes are a combination of regions from different genetic origin. The classic strains are mostly of *M. m. domesticus* origin (86 to 96% of the genome) with only small contributions from the other two subspecies[338]. Therefore, the genomes from B6 and 129 animals are closely related; these contain only around 4.4 million SNPs and 800 thousand small indels[321]. As a comparison, any two humans differ, on average, at around two to three million basepairs, considering SNPs only[339]. In contrast, CAST animals are a wild-derived strain, that pertains to the *M. m. castaneus* subdivision. As such, it is a lot more divergent from the inbred classic laboratory strains, and contains more than four times the amount of variation[321]. Despite the disparate divergence of these strains, the OR expression levels for the whole repertoire are remarkably dissimilar between all, with up to 50% of all receptor genes significantly differentially expressed. This indicates that genetic variation has a very significant effect on the regulation of the final distribution of the different OSN types (Figure 6.1A-B).

Analysis of the distribution of genetic variation in the mouse genome has revealed that OR genes have slightly more variation than the average gene[271] and they tend to be enriched in regions of copy number variation both in humans[255] and mice[340]. These two characteristics reflect the evolutionary dynamics of the OR gene family. The olfactory system has evolved to discriminate a large catalog of molecules by diversifying the repertoire of receptors available for detection. The ability to sense a larger number of odorants increases the amount of information an animal can gather from their surroundings, and provides a reservoir of detectors to adapt to novel environments. As such, diversification of the OR repertoire should be beneficial. Analysis of introgression events between different mouse species lends support for this hypothesis. Introgression events between the house mouse (*M. m. domesticus*) and the Algerian mouse (*Mus spretus*) are common in the wild, but most hybridisation events tend to be removed by drift and selection[341]. An analysis of the genomes from diverse wild-caught mice identified some regions where hybridisation occurred between the two species. Interestingly, the intro-

gressed regions that have prevailed in the genomes of several individuals, are enriched in OR genes, suggesting that their maintenance is beneficial to the animals[341].

Based on the above, genetic variation to diversify the OR repertoire is beneficial. Several examples have shown that single amino acid changes are able to shift the binding specificity of an OR[224] and, often, they result in the innervation of separate glomeruli[85]. In contrast to inbred mice, wild animals will possess up to twice as many alleles for the OR gene catalogue. Thus, individuals heterozygous at some OR loci will likely possess increased detection and information processing capabilities. Furthermore, genomic variation will also affect the non-coding portions of the OR genes, their promoters and regulatory elements. Alteration of transcription factor binding sites has been shown to impact the final number of OSNs that express the affected allele[188]. Thus, the accumulation of non-coding genetic variation will have an effect on the proportion of the neuronal population that is taken by each particular allele (Figure 6.1C). Hence, the combination of coding variation that alters the detection properties of the receptors, along with non-coding variants that modify the number of OSNs that express each OR, will ultimately produce a unique repertoire of ORs with a specific OSN distribution, which in turn will impact olfactory sensing.

In support of this, analysis of the 1000 Genomes Project data has revealed that any two individuals differ in around 30% of their OR genes, either by possessing differing sets of segregating pseudogenes or by coding variation that has an impact on the response profile of the receptor to its ligands[241]. Further, several examples exist on the effects of genetic variation in OR genes and differences in perception[241, 266–268, 270]. A recent study tested the variability in human perception of a set of odorants, based on several descriptors. While the gross perception based on pleasantness was very similar between individuals, it was highly specific when detailed descriptors were used. Furthermore, the perceptual profiles were highly variable between individuals; so much so, that the authors proposed that with enough odorants and descriptors, it would be possible to create an *olfactory fingerprint* for every person. What's more, the similarity of two perceptual fingerprints was correlated with the similarity of their HLA profiles (human leukocyte antigen system, analogous to the major histocompatibility complex in animals), suggesting that the olfactory fingerprint might be capturing genetic information[342]. Thus, these data suggest that human perception is indeed highly variable, as is the OR profile of each individual.

6.4 Plastic control of OSN diversity.

Little information is available on how differences in the number of OSNs expressing a particular OR gene affect detection and/or perception of a particular stimulus. An interesting study generated a mouse with a "monoclonal nose", where over 95% of all the OSNs expressed the M71 receptor. Correspondingly, the rest of the OR repertoire was dramatically reduced. The small number of neurones expressing other receptor types still innervated particular glomeruli, though these were co-innervated by M71 axons. Not surprisingly, EOG recordings upon exposure to acetophenone (the ligand of M71) were greatly increased, and responses to other odorants were diminished. Similarly, acetophenone elicited widespread glomerular activation while other odorants did not elicit detectable responses. Despite the low number of OSNs expressing most receptors, the mice were able to detect and discriminate between different odorants and even between enantiomer pairs. However, their ability to differentiate mixtures of enantiomers was greatly impaired[343]. Thus, these data suggests that a low number of OSNs expressing a given receptor are sufficient to bind odorants and transmit the information, which can be used for olfactory learning tasks. However, the discrimination capacity is greatly weakened, perhaps because the glomerular activation is not strong enough to allow differentiation between similar patterns. Unfortunately, the authors did not test the detection threshold of these animals to common odorants. Therefore, it is unclear whether these animals also have reduced sensitivity; it could be that they are able to detect a ligand only when it is present at high concentrations.

It is tempting to speculate that animals with varying proportions of each receptor type will have different capabilities to detect and discriminate differing sets of odorants. If so, the ability to tune the proportion of OSNs devoted to the recognition of important odorants would be greatly beneficial, especially if the starting abundance dictated by the genetic background is low. Several studies have shown that odorant stimulation increases the life-span of the OSNs that are activated and, with time, these OSN types become enriched in the MOE[286, 288, 292]. Consistently, I have found that the intermittent exposure of animals to either a cocktail of four different odorants, or subsets of these, results in the differential expression of specific OR genes (Figure 6.1D). Interestingly, no changes could be detected when the odorants were present 24 hours a day. Presumably, an odorant that is always part of an animal's environment is non-informative and thus

it would not be advantageous to devote any more OSNs to its continuous detection. At the OSN level, adaptation stops OSNs from responding to sustained stimulation. Based on my results, it seems likely that constant stimulation also blocks enhanced survival.

In the acutely exposed animals the changes in OR expression observed increased with time, which argues in favour of a survival-mediated mechanism. Moreover, it is difficult to imagine a plausible mechanism through which an odorant can influence the choice of its cognate OR during neurogenesis. However, in all experiments, a subset of genes were also consistently downregulated, suggesting a decrease in OSN number. Though unexpected at first glance, the adult MOE maintains a fine balance of the total number of OSNs[22]; therefore, to increase the frequency of some OSN types it may be necessary to decrease others. If all OR genes were to decrease equally, it is likely that the changes would be small enough not to be detected by expression profiling. However, a more parsimonious scenario is one where the receptors expressed in overlapping regions with those that increase frequency are the ones affected, while the rest of the repertoire remains unchanged. This has been observed in a mouse where the coding sequence of *MOR28* was removed; only ORs expressed in the same zone were able to populate the OSNs initially devoted to express MOR28[207]. Thus, it is not unreasonable to assume that the increase in a particular OSN type would take the space of OSNs from the same region. Still, it is likely that many ORs are expressed in the region overlapping that of an activated OR; hence the downregulation of each should be small and it might require longer times to reach a differential that is detectable by RNAseq. Indeed, the number of downregulated ORs is much larger (44.4% of the total) in the animals that were exposed for 24 weeks, than in those that were exposed for only 10 weeks (6.7%, 25% and 33.3% for the groups exposed to R-carvone, heptanal or both).

Recently, a study reported that short-term exposure (5 hours) of animals to particular odorants results in the downregulation of the ORs that respond to them, at the mRNA level; for some OR genes, these changes could be observed as soon as 30 minutes after the start of the exposure[344], but no shorter times were tested. Therefore, it is not clear whether this could be occurring in the acutely exposed animals. The authors performed a comprehensive analysis by RNAseq of animals exposed to acetophenone and generated a list of downregulated OR genes. These were further shown to colocalise with phosphorylated ribosomal protein S6 (pS6), which is a marker of activated OSNs[344]. In parallel, an independent group used the presence of pS6 to capture the OSNs that were activated upon stimulation, and then performed RNAseq to identify the ORs expressed; these were then validated in a heterologous cell system[345]. Together, the lists of

receptors identified by these groups, overlap with four different OR genes that were significantly DE in the animals exposed to the mix of four odorants (24 weeks), that contained acetophenone. Interestingly, all four receptors were downregulated in the RNAseq data.

Based on the above, it seems that at least some of the downregulated OR genes are activated by acetophenone. Thus, it could be that only neurones expressing some OR genes are able to modulate their life-span or that the increase of some OSN types is not big enough to be detected by RNAseq; instead, the temporary downregulation from exposure events close to the time of tissue collection could be identified as overall downregulation. Further, it could also be possible that different mechanisms operate depending on the affinity of each receptor for a given ligand. To better understand the dynamics of the changes observed, I have established a collaboration with Casey Trimmer and Joel Mainland (Monell Chemical Senses Center) to test some of the DE ORs in an *in vitro* response assay in heterologous cells. Preliminary results indicate that three out of five DE ORs tested indeed respond to the mix of four odorants used as stimulus (data not shown), but more systematic and thorough tests are being carried out at present.

Conflicting data is available on the effect of odorant exposure on OR expression and OSN number. Whereas several studies have concluded that OSN activation leads to increased life-span which, with time, should increase OSN number[286, 288, 292], others have proposed that olfactory stimulation results in a reduced number of the activated OSNs[293, 294]. In some cases, these changes have been shown to be specific to a particular OR, whereas other receptors remain unchanged[294]. Thus, while the analysis of particular OR-ligand pairs reveal interesting phenomena, the observations cannot be generalised. In this respect, my data provides the first comprehensive study of the response of the complete OR repertoire to a particular olfactory stimulation paradigm.

6.5 Functional impact of differences in OSN number.

As mentioned previously, it is not clear what is the functional consequence of altering the number of OSNs that express a particular OR. One hypothesis is that a greater number of detectors would result in enhanced sensitivity towards the odorants that are recognised with high affinity. To directly test this, I have created a transgenic mouse

line –in a B6 genetic background– where the CDS of *Olfr1507* (the most abundant OR in this strain) has been replaced by that of *Olfr2*. For this I utilised CRISPR-Cas9 technology. I created a vector for homologous recombination (HR) that contained the coding sequence of *Olfr2*, flanked by 1kb homology arms matching the *Olfr1507* locus. The vector was microinjected into B6 embryos, along with two guideRNA molecules that produce double-strand cuts in the intended site of HR. These embryos were then allowed to develop to term in foster mothers.

Olfr1507 is expressed 35 times more abundantly than *Olfr2*. Since the abundance of a particular OR gene is controlled by the genetic architecture in *cis*, I expect to greatly increase the expression –and therefore the number of OSNs– of *Olfr2* in these animals. The response profile of *Olfr2* (better known as *I7*) has been very well characterised and, thus, these transgenic animals will provide an opportunity to assess the impact of increasing the cell number of a given OR on odour detection.

References

[1] Ibarra-Soria, X., Levitin, M. O. & Logan, D. W. The genomic basis of vomeronasal-mediated behaviour. *Mamm Genome* **25**, 75–86 (2014).

[2] Mombaerts, P. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci* **5**, 263–278 (2004).

[3] Breer, H., Fleischer, J. & Strotmann, J. The sense of smell: multiple olfactory subsystems. *Cellular and Molecular Life Sciences CMLS* **63**, 1465–1475 (2006).

[4] Brennan, P. A. & Zufall, F. Pheromonal communication in vertebrates. *Nature* **444**, 308–315 (2006).

[5] Touhara, K. & Vosshall, L. B. Sensing odorants and pheromones with chemosensory receptors. *Annual Review of Physiology* **71**, 307–332 (2009).

[6] Firestein, S. How the olfactory system makes sense of scents. *Nature* **413**, 211–218 (2001).

[7] Bushdid, C., Magnasco, M. O., Vosshall, L. B. & Keller, A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **343**, 1370–1372 (2014).

[8] Gerkin, R. C. & Castro, J. B. The number of olfactory stimuli that humans can discriminate is still unknown. *Elife* **4** (2015).

[9] Meister, M. On the dimensionality of odor space. *Elife* **4** (2015).

[10] Wyatt, T. D. *Pheromones and Animal Behaviour: Communication by Smell and Taste* (Cambridge University Press, 2003).

[11] Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nat Rev Neurosci* **4**, 551–562 (2003).

[12] Spehr, M. *et al.* Parallel processing of social signals by the mammalian main and accessory olfactory systems. *Cell Mol Life Sci* **63**, 1476–1484 (2006).

[13] Morrison, E. E. & Costanzo, R. M. Morphology of olfactory epithelium in humans and other vertebrates. *Microscopy Research and Technique* **23**, 49–61 (1992).

[14] Hopkins, A. E. The olfactory receptors in vertebrates. *The Journal of Comparative Neurology* **41**, 253–289 (1926).

[15] Frisch, D. Ultrastructure of mouse olfactory mucosa. *American Journal of Anatomy* **121**, 87–119 (1967).

[16] Lancet, D. Vertebrate olfactory reception. *Annual Review of Neuroscience* **9**, 329–355 (1986).

[17] Moulton, D. G. & Beidler, L. M. Structure and function in the peripheral olfactory system. *Physiological Reviews* **47**, 1–52 (1967).

[18] Strotmann, J. *et al.* Olfactory neurones expressing distinct odorant receptor subtypes are spatially segregated in the nasal neuroepithelium. *Cell and Tissue Research* **276**, 429–438 (1994).

[19] Caggiano, M., Kauer, J. S. & Hunter, D. D. Globose basal cells are neuronal progenitors in the olfactory epithelium: A lineage analysis using a replication-incompetent retrovirus. *Neuron* **13**, 339 – 352 (1994).

[20] Leung, C. T., Coulombe, P. A. & Reed, R. R. Contribution of olfactory neural stem cells to tissue maintenance and regeneration. *Nature neuroscience* **10**, 720–726 (2007).

[21] Calof, A. L. & Chikaraishi, D. M. Analysis of neurogenesis in a mammalian neuroepithelium: Proliferation and differentiation of an olfactory neuron precursor in vitro. *Neuron* **3**, 115 – 127 (1989).

[22] Mackay-Sim, A. Stem cells and their niche in the adult olfactory mucosa. *Arch Ital Biol* **148**, 47–58 (2010).

[23] Kawagishi, K. *et al.* Stereological quantification of olfactory receptor neurons in mice. *Neuroscience* **272**, 29 – 33 (2014).

[24] Nagashima, A. & Touhara, K. Enzymatic conversion of odorants in nasal mucus affects olfactory glomerular activation patterns and odor perception. *J Neurosci* **30**, 16391–16398 (2010).

[25] Mori, K. & Sakano, H. How is the olfactory map formed and interpreted in the mammalian brain? *Annu Rev Neurosci* **34**, 467–499 (2011).

[26] Hinds, J. W., Hinds, P. L. & McNelly, N. A. An autoradiographic study of the mouse olfactory epithelium: evidence for long-lived receptors. *Anat Rec* **210**, 375–383 (1984).

[27] Graziadei, P. P. & Graziadei, G. A. Neurogenesis and neuron regeneration in the olfactory system of mammals. i. morphological aspects of differentiation and structural organization of the olfactory sensory neurons. *J Neurocytol* **8**, 1–18 (1979).

[28] Mackay-Sim, A. & Kittel, P. Cell dynamics in the adult mouse olfactory epithelium: a quantitative autoradiographic study. *J Neurosci* **11**, 979–984 (1991).

[29] Mackay-Sim, A. & Kittel, P. W. On the life span of olfactory receptor neurons. *Eur J Neurosci* **3**, 209–215 (1991).

[30] Kondo, K. *et al.* Age-related changes in cell dynamics of the postnatal mouse olfactory neuroepithelium: cell proliferation, neuronal differentiation, and cell death. *J Comp Neurol* **518**, 1962–1975 (2010).

[31] Diaz, D. *et al.* The olfactory system as a puzzle: playing with its pieces. *Anat Rec (Hoboken)* **296**, 1383–1400 (2013).

[32] Brann, J. H. & Firestein, S. A lifetime of neurogenesis in the olfactory system. *Frontiers in Neuroscience* **8** (2014).

[33] Murray, R. C. & Calof, A. L. Neuronal regeneration: lessons from the olfactory system. *Semin Cell Dev Biol* **10**, 421–431 (1999).

[34] Pace, U., Hanski, E., Salomon, Y. & Lancet, D. Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature* **316**, 255–258 (1985).

[35] Sklar, P. B., Anholt, R. R. & Snyder, S. H. The odorant-sensitive adenylate cyclase of olfactory receptor cells. differential stimulation by distinct classes of odorants. *Journal of Biological Chemistry* **261**, 15538–43 (1986).

[36] Boekhoff, I., Tareilus, E., Strotmann, J. & Breer, H. Rapid activation of alternative second messenger pathways in olfactory cilia from rats by different odorants. *The EMBO journal* **9**, 2453–2458 (1990).

[37] Breer, H., Boekhoff, I. & Tareilus, E. Rapid kinetics of second messenger formation in olfactory transduction. *Nature* **345**, 65–68 (1990).

[38] Nakamura, T. & Gold, G. H. A cyclic nucleotide-gated conductance in olfactory receptor cilia. *Nature* **325**, 442–444 (1987).

[39] Dhallan, R. S., Yau, K.-W., Schrader, K. A. & Reed, R. R. Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons. *Nature* **347**, 184–187 (1990).

[40] Kleene, S. & Gesteland, R. Calcium-activated chloride conductance in frog olfactory cilia. *The Journal of Neuroscience* **11**, 3624–3629 (1991).

[41] Kleene, S. J. Origin of the chloride current in olfactory transduction. *Neuron* **11**, 123–132 (1993).

[42] Lowe, G. & Gold, G. H. Nonlinear amplification by calcium-dependent chloride channels in olfactory receptor cells. *Nature* **366**, 283–286 (1993).

[43] Kurahashi, T. & Yau, K. W. Co-existence of cationic and chloride components in odorant-induced current of vertebrate olfactory receptor cells. *Nature* **363**, 71–74 (1993).

[44] Jones, D. & Reed, R. Golf: an olfactory neuron specific-g protein involved in odorant signal transduction. *Science* **244**, 790–795 (1989).

[45] Wong, S. T. *et al.* Disruption of the type {III} adenylyl cyclase gene leads to peripheral and behavioral anosmia in transgenic mice. *Neuron* **27**, 487 – 497 (2000).

[46] Stephan, A. B. *et al.* Ano2 is the cilial calcium-activated chloride channel that may mediate olfactory amplification. *Proceedings of the National Academy of Sciences* **106**, 11776–11781 (2009).

[47] Kaupp, U. B. Olfactory signalling in vertebrates and insects: differences and commonalities. *Nat Rev Neurosci* **11**, 188–200 (2010).

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

[49] Baker, H. *et al.* Targeted deletion of a cyclic nucleotide-gated channel subunit (ocnc1): Biochemical and morphological consequences in adult mice. *The Journal of Neuroscience* **19**, 9313–9321 (1999).

[50] Belluscio, L., Gold, G. H., Nemes, A. & Axel, R. Mice deficient in golf are anosmic. *Neuron* **20**, 69–81 (1998).

[51] Logan, D. W. *et al.* Learned recognition of maternal signature odors mediates the first suckling episode in mice. *Curr Biol* **22**, 1998–2007 (2012).

[52] Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **65**, 175 – 187 (1991).

[53] Imai, T. & Sakano, H. Odorant receptor gene choice and axonal projection in the mouse olfactory system. *Results Probl Cell Differ* **47**, 57–75 (2009).

[54] Glusman, G., Yanai, I., Rubin, I. & Lancet, D. The complete human olfactory subgenome. *Genome Research* **11**, 685–702 (2001).

[55] Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nature neuroscience* **5**, 124–133 (2002).

[56] Young, J. M. *et al.* Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Human Molecular Genetics* **11**, 535–546 (2002).

[57] Godfrey, P. A., Malnic, B. & Buck, L. B. The mouse olfactory receptor gene family. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2156–2161 (2004).

[58] Zhang, X., Zhang, X. & Firestein, S. Comparative genomics of odorant and pheromone receptor genes in rodents. *Genomics* **89**, 441 – 450 (2007).

[59] Freitag, J., Krieger, J., Strotmann, J. & Breer, H. Two classes of olfactory receptors in xenopus laevis. *Neuron* **15**, 1383 – 1392 (1995).

[60] Mombaerts, P. Axonal wiring in the mouse olfactory system. *Annu Rev Cell Dev Biol* **22**, 713–737 (2006).

[61] Sullivan, S. L., Adamson, M. C., Ressler, K. J., Kozak, C. A. & Buck, L. B. The chromosomal distribution of mouse odorant receptor genes. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 884–888 (1996).

[62] Young, J. M. *et al.* Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels. *Genome Biol* **4**, R71 (2003).

[63] Mombaerts, P. Molecular biology of odorant receptors in vertebrates. *Annual Review of Neuroscience* **22**, 487–509 (1999).

[64] Liu, A. H., Zhang, X., Stolovitzky, G. A., Califano, A. & Firestein, S. J. Motif-based construction of a functional map for mammalian olfactory receptors. *Genomics* **81**, 443 – 456 (2003).

[65] Young, J. M. & Trask, B. J. V2r gene families degenerated in primates, dog and cow, but expanded in opossum. *Trends Genet* **23**, 212–5 (2007).

[66] Ressler, K. J., Sullivan, S. L. & Buck, L. B. A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**, 597 – 609 (1993).

[67] Vassar, R., Ngai, J. & Axel, R. Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* **74**, 309 – 318 (1993).

[68] Kubick, S., Strotmann, J., Andreini, I. & Breer, H. Subfamily of olfactory receptors characterized by unique structural features and expression patterns. *Journal of Neurochemistry* **69**, 465–475 (1997).

[69] Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713 – 723 (1999).

[70] Strotmann, J. *et al.* Local permutations in the glomerular array of the mouse olfactory bulb. *The Journal of Neuroscience* **20**, 6927–6938 (2000).

[71] Serizawa, S. *et al.* Mutually exclusive expression of odorant receptor transgenes. *Nat Neurosci* **3**, 687–693 (2000).

[72] Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823 – 834 (1994).

[73] Ishii, T. *et al.* Monoallelic expression of the odourant receptor gene and axonal projection of olfactory sensory neurones. *Genes Cells* **6**, 71–78 (2001).

[74] Mombaerts, P. Odorant receptor gene choice in olfactory sensory neurons: the one receptor–one neuron hypothesis revisited. *Current Opinion in Neurobiology* **14**, 31 – 36 (2004).

[75] Strotmann, J., Wanner, L., Helfrich, T., Beck, A. & Breer, H. Rostro-caudal patterning of receptor-expressing olfactory neurones in the rat nasal cavity. *Cell and Tissue Research* **278**, 11–20 (1994).

[76] Tsuboi, A., Miyazaki, T., Imai, T. & Sakano, H. Olfactory sensory neurons expressing class i odorant receptors converge their axons on an antero-dorsal domain of the olfactory bulb in the mouse. *European Journal of Neuroscience* **23**, 1436–1444 (2006).

[77] Scott, J. W. & Brierley, T. A functional map in rat olfactory epithelium. *Chem Senses* **24**, 679–690 (1999).

[78] Miyamichi, K., Serizawa, S., Kimura, H. M. & Sakano, H. Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *The Journal of Neuroscience* **25**, 3586–3592 (2005).

[79] Ressler, K. J., Sullivan, S. L. & Buck, L. B. Information coding in the olfactory system: Evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245 – 1255 (1994).

[80] Feinstein, P. & Mombaerts, P. A contextual model for axonal sorting into glomeruli in the mouse olfactory system. *Cell* **117**, 817–831 (2004).

[81] Vassar, R. *et al.* Topographic organization of sensory projections to the olfactory bulb. *Cell* **79**, 981–991 (1994).

[82] Schaefer, M. L., Finger, T. E. & Restrepo, D. Variability of position of the p2 glomerulus within a map of the mouse olfactory bulb. *J Comp Neurol* **436**, 351–362 (2001).

[83] DeMaria, S. & Ngai, J. The cell biology of smell. *J Cell Biol* **191**, 443–452 (2010).

[84] Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675 – 686 (1996).

[85] Feinstein, P., Bozza, T., Rodriguez, I., Vassalli, A. & Mombaerts, P. Axon guidance of mouse olfactory sensory neurons by odorant receptors and the beta2 adrenergic receptor. *Cell* **117**, 833–846 (2004).

[86] Wang, H. W., Wysocki, C. J. & Gold, G. H. Induction of olfactory receptor sensitivity in mice. *Science* **260**, 998–1000 (1993).

[87] Zhang, J., Huang, G., Dewan, A., Feinstein, P. & Bozza, T. Uncoupling stimulus specificity and glomerular position in the mouse olfactory system. *Mol Cell Neurosci* **51**, 79–88 (2012).

[88] Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M. & Mombaerts, P. Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb. *Neuron* **35**, 681–696 (2002).

[89] Cho, J. H., Lépine, M., Andrews, W., Parnavelas, J. & Cloutier, J.-F. Requirement for slit-1 and robo-2 in zonal segregation of olfactory sensory neuron axons in the main olfactory bulb. *The Journal of Neuroscience* **27**, 9094–9104 (2007).

[90] Takeuchi, H. *et al.* Sequential arrival and graded secretion of sema3f by olfactory neuron axons specify map topography at the bulb. *Cell* **141**, 1056 – 1067 (2010).

[91] Bozza, T. *et al.* Mapping of class i and class ii odorant receptors to glomerular domains by two distinct types of olfactory sensory neurons in the mouse. *Neuron* **61**, 220–233 (2009).

[92] Imai, T., Suzuki, M. & Sakano, H. Odorant receptor derived camp signals direct axonal targeting. *Science* **314**, 657–661 (2006).

[93] Connelly, T., Savigner, A. & Ma, M. Spontaneous and sensory-evoked activity in mouse olfactory sensory neurons with defined odorant receptors. *J Neurophysiol* **110**, 55–62 (2013).

[94] Col, J. A. D., Matsuo, T., Storm, D. R. & Rodriguez, I. Adenylyl cyclase-dependent axonal targeting in the olfactory system. *Development* **134**, 2481–2489 (2007).

[95] Imai, T. *et al.* Pre-target axon sorting establishes the neural map topography. *Science* **325**, 585–590 (2009).

[96] Zou, D.-J. *et al.* Postnatal refinement of peripheral olfactory projections. *Science* **304**, 1976–1979 (2004).

[97] Serizawa, S. *et al.* A neuronal identity code for the odorant receptor-specific and activity-dependent axon sorting. *Cell* **127**, 1057 – 1069 (2006).

[98] Liberles, S. D. & Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**, 645–650 (2006).

[99] Johnson, M. A. *et al.* Neurons expressing trace amine-associated receptors project to discrete glomeruli and constitute an olfactory subsystem. *Proceedings of the National Academy of Sciences* **109**, 13410–13415 (2012).

[100] Dewan, A., Pacifico, R., Zhan, R., Rinberg, D. & Bozza, T. Non-redundant coding of aversive odours in the main olfactory pathway. *Nature* **497**, 486–489 (2013).

[101] Pacifico, R., Dewan, A., Cawley, D., Guo, C. & Bozza, T. An olfactory subsystem that mediates high-sensitivity detection of volatile amines. *Cell Reports* **2**, 76 – 88 (2012).

[102] Liberles, S. D. Trace amine-associated receptors: ligands, neural circuits, and behaviors. *Curr Opin Neurobiol* **34C**, 1–7 (2015).

[103] Ferrero, D. M. *et al.* Detection and avoidance of a carnivore odor by prey. *Proceedings of the National Academy of Sciences* **108**, 11235–11240 (2011).

[104] Zhang, J., Pacifico, R., Cawley, D., Feinstein, P. & Bozza, T. Ultrasensitive detection of amines by a trace amine-associated receptor. *The Journal of Neuroscience* **33**, 3228–3239 (2013).

[105] Li, Q. *et al.* Synchronous evolution of an odor biosynthesis pathway and behavioral response. *Current Biology* **23**, 11 – 20 (2013).

[106] Fulle, H. J. *et al.* A receptor guanylyl cyclase expressed specifically in olfactory sensory neurons. *Proc Natl Acad Sci U S A* **92**, 3571–3575 (1995).

[107] Juilfs, D. M. *et al.* A subset of olfactory neurons that selectively express cgmp-stimulated phosphodiesterase (pde2) and guanylyl cyclase-d define a unique olfactory signal transduction pathway. *Proc Natl Acad Sci U S A* **94**, 3388–3395 (1997).

[108] Meyer, M. R., Angele, A., Kremmer, E., Kaupp, U. B. & Muller, F. A cgmp-signaling pathway in a subset of olfactory sensory neurons. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 10595–10600 (2000).

[109] Hu, J. *et al.* Detection of near-atmospheric concentrations of co2 by an olfactory subsystem in the mouse. *Science* **317**, 953–957 (2007).

[110] Sun, L. *et al.* Guanylyl cyclase-d in the olfactory co2 neurons is activated by bicarbonate. *Proceedings of the National Academy of Sciences* **106**, 2041–2046 (2009).

[111] Galef, B. G. & Wigmore, S. W. Transfer of information concerning distant foods: A laboratory investigation of the 'information-centre' hypothesis. *Animal Behaviour* **31**, 748 – 758 (1983).

[112] Munger, S. D. *et al.* An olfactory subsystem that detects carbon disulfide and mediates food-related social learning. *Current biology : CB* **20**, 1438–1444 (2010).

[113] Leinders-Zufall, T. *et al.* Contribution of the receptor guanylyl cyclase gc-d to chemosensory function in the olfactory epithelium. *Proceedings of the National Academy of Sciences* **104**, 14507–14512 (2007).

[114] Arakawa, H., Kelliher, K. R., Zufall, F. & Munger, S. D. The receptor guanylyl cyclase type d (gc-d) ligand uroguanylin promotes the acquisition of food preferences in mice. *Chemical Senses* **38**, 391–397 (2013).

[115] Døving, K. B. & Trotier, D. Structure and function of the vomeronasal organ. *The Journal of Experimental Biology* **201**, 2913–2925 (1998).

[116] Keverne, E. B. The vomeronasal organ. *Science* **286**, 716–720 (1999).

[117] Berghard, A. & Buck, L. Sensory transduction in vomeronasal neurons: evidence for g alpha o, g alpha i2, and adenylyl cyclase ii as major components of a pheromone signaling cascade. *The Journal of Neuroscience* **16**, 909–918 (1996).

[118] Norlin, E., Gussing, F. & Berghard, A. Vomeronasal phenotype and behavioral alterations in g(alpha)i2 mutant mice. *Current Biology* **13**, 1214 – 1219 (2003).

[119] Rudolph, U. *et al.* Ulcerative colitis and adenocarcinoma of the colon in g[alpha]i2-deficient mice. *Nat Genet* **10**, 143–150 (1995).

[120] Chamero, P. *et al.* G protein g(alpha)o is essential for vomeronasal function and aggressive behavior in mice. *Proc Natl Acad Sci U S A* **108**, 12898–903 (2011).

[121] Liman, E. R., Corey, D. P. & Dulac, C. Trp2: A candidate transduction channel for mammalian pheromone sensory signaling. *Proceedings of the National Academy of Sciences* **96**, 5791–5796 (1999).

[122] Menco, B. P. M., Carr, V. M., Ezeh, P. I., Liman, E. R. & Yankova, M. P. Ultrastructural localization of g-proteins and the channel protein trp2 to microvilli of rat vomeronasal receptor cells. *The Journal of Comparative Neurology* **438**, 468–489 (2001).

[123] Leypold, B. G. *et al.* Altered sexual and social behaviors in trp2 mutant mice. *Proceedings of the National Academy of Sciences* **99**, 6376–6381 (2002).

[124] Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. Loss of sex discrimination and male-male aggression in mice deficient for trp2. *Science* **295**, 1493–1500 (2002).

[125] Yang, C. & Delay, R. J. Calcium-activated chloride current amplifies the response to urine in mouse vomeronasal sensory neurons. *The Journal of General Physiology* **135**, 3–13 (2010).

[126] Kim, S., Ma, L. & Yu, C. R. Requirement of calcium-activated chloride channels in the activation of mouse vomeronasal neurons. *Nat Commun* **2**, 365 (2011).

[127] Hasen, N. S. & Gammie, S. C. Trpc2-deficient lactating mice exhibit altered brain and behavioral responses to bedding stimuli. *Behavioural Brain Research* **217**, 347 – 353 (2011).

[128] Kimchi, T., Xu, J. & Dulac, C. A functional circuit underlying male sexual behaviour in the female mouse brain. *Nature* **448**, 1009–1014 (2007).

[129] Ben-Shaul, Y., Katz, L. C., Mooney, R. & Dulac, C. In vivo vomeronasal stimulation reveals sensory encoding of conspecific and allospecific cues by the mouse accessory olfactory bulb. *Proceedings of the National Academy of Sciences* **107**, 5172–5177 (2010).

[130] Papes, F., Logan, D. W. & Stowers, L. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141**, 692–703 (2010).

[131] Omura, M. & Mombaerts, P. Trpc2-expressing sensory neurons in the main olfactory epithelium of the mouse. *Cell Reports* **8**, 583–595 (2014).

[132] Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).

[133] Herrada, G. & Dulac, C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**, 763–73 (1997).

[134] Matsunami, H. & Buck, L. B. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**, 775–84 (1997).

[135] Ryba, N. J. & Tirindelli, R. A new multigene family of putative pheromone receptors. *Neuron* **19**, 371–9 (1997).

[136] Young, J. M., Massa, H. F., Hsu, L. & Trask, B. J. Extreme variability among mammalian v1r gene families. *Genome Res* **20**, 10–8 (2010).

[137] Rodriguez, I., Del Punta, K., Rothman, A., Ishii, T. & Mombaerts, P. Multiple new and isolated families within the mouse superfamily of v1r vomeronasal receptors. *Nature neuroscience* **5**, 134–140 (2002).

[138] Zhang, X., Rodriguez, I., Mombaerts, P. & Firestein, S. Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics* **83**, 802–11 (2004).

[139] Leinders-Zufall, T. *et al.* Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).

[140] Boschat, C. *et al.* Pheromone detection mediated by a v1r vomeronasal receptor. *Nat Neurosci* **5**, 1261–2 (2002).

[141] Isogai, Y. *et al.* Molecular organization of vomeronasal chemoreception. *Nature* **478**, 241–245 (2011).

[142] Nodari, F. *et al.* Sulfated steroids as natural ligands of mouse pheromone-sensing neurons. *J Neurosci* **28**, 6407–18 (2008).

[143] Del Punta, K. *et al.* Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature* **419**, 70–4 (2002).

[144] Dulac, C. & Wagner, S. Genetic analysis of brain circuits underlying pheromone signaling. *Annual Review of Genetics* **40**, 449–467 (2006).

[145] Chamero, P., Leinders-Zufall, T. & Zufall, F. From genes to social communication: molecular sensing by the vomeronasal organ. *Trends Neurosci* **35**, 597–606 (2012).

[146] Yang, H., Shi, P., Zhang, Y. P. & Zhang, J. Composition and evolution of the v2r vomeronasal receptor gene repertoire in mice and rats. *Genomics* **86**, 306–15 (2005).

[147] Ishii, T. & Mombaerts, P. Coordinated coexpression of two vomeronasal receptor v2r genes per neuron in the mouse. *Mol Cell Neurosci* **46**, 397–408 (2011).

[148] Martini, S., Silvotti, L., Shirazi, A., Ryba, N. J. & Tirindelli, R. Co-expression of putative pheromone receptors in the sensory neurons of the vomeronasal organ. *J Neurosci* **21**, 843–8 (2001).

[149] Silvotti, L., Moiani, A., Gatti, R. & Tirindelli, R. Combinatorial co-expression of pheromone receptors, v2rs. *J Neurochem* **103**, 1753–63 (2007).

[150] Ishii, T., Hirota, J. & Mombaerts, P. Combinatorial coexpression of neural and immune multigene families in mouse vomeronasal sensory neurons. *Curr Biol* **13**, 394–400 (2003).

[151] Loconto, J. *et al.* Functional expression of murine v2r pheromone receptors involves selective association with the m10 and m1 families of mhc class ib molecules. *Cell* **112**, 607–18 (2003).

[152] Leinders-Zufall, T. *et al.* Mhc class i peptides as chemosensory signals in the vomeronasal organ. *Science* **306**, 1033–1037 (2004).

[153] Leinders-Zufall, T., Ishii, T., Mombaerts, P., Zufall, F. & Boehm, T. Structural requirements for the activation of vomeronasal sensory neurons by mhc peptides. *Nature neuroscience* **12**, 1551–1558 (2009).

[154] Bruce, H. An exteroceptive block to pregnancy in the mouse. *Nature* **184** (1959).

[155] Wyatt, T. D. Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. *Journal of Comparative Physiology A* **196**, 685–700 (2010).

[156] Chamero, P. *et al.* Identification of protein pheromones that promote aggressive behaviour. *Nature* **450**, 899–902 (2007).

[157] Haga, S. *et al.* The male mouse pheromone esp1 enhances female sexual receptive behaviour through a specific vomeronasal receptor. *Nature* **466**, 118–22 (2010).

[158] Hurst, J. L. *et al.* Individual recognition in mice mediated by major urinary proteins. *Nature* **414**, 631–4 (2001).

[159] Roberts, S. *et al.* Darcin: a male pheromone that stimulates female memory and sexual attraction to an individual male's odour. *BMC biology* **8**, 75 (2010).

[160] Liberles, S. D. *et al.* Formyl peptide receptors are candidate chemosensory receptors in the vomeronasal organ. *Proc Natl Acad Sci U S A* **106**, 9842–7 (2009).

[161] Riviere, S., Challet, L., Fluegge, D., Spehr, M. & Rodriguez, I. Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature* **459**, 574–7 (2009).

[162] Bufe, B., Schumann, T. & Zufall, F. Formyl peptide receptors from immune and vomeronasal system exhibit distinct agonist properties. *Journal of Biological Chemistry* **287**, 33644–33655 (2012).

[163] Yang, H. & Shi, P. Molecular and evolutionary analyses of formyl peptide receptors suggest the absence of vno-specific fprs in primates. *J Genet Genomics* **37**, 771–8 (2010).

[164] Ma, M. *et al.* Olfactory signal transduction in the mouse septal organ. *The Journal of Neuroscience* **23**, 317–324 (2003).

[165] Giannetti, N., Saucier, D. & Astic, L. Organization of the septal organ projection to the main olfactory bulb in adult and newborn rats. *J Comp Neurol* **323**, 288–298 (1992).

[166] Kaluza, J. F., Gussing, F., Bohm, S., Breer, H. & Strotmann, J. Olfactory receptors in the mouse septal organ. *Journal of Neuroscience Research* **76**, 442–452 (2004).

[167] Tian, H. & Ma, M. Molecular organization of the olfactory septal organ. *The Journal of Neuroscience* **24**, 8383–8390 (2004).

[168] Ma, M. Encoding olfactory signals via multiple chemosensory systems. *Crit Rev Biochem Mol Biol* **42**, 463–480 (2007).

[169] Levai, O. & Strotmann, J. Projection pattern of nerve fibers from the septal organ: Dii-tracing studies with transgenic omp mice. *Histochemistry and Cell Biology* **120**, 483–492 (2003).

[170] Storan, M. J. & Key, B. Septal organ of gruneberg is part of the olfactory system. *J Comp Neurol* **494**, 834–844 (2006).

[171] Fuss, S. H., Omura, M. & Mombaerts, P. The grueneberg ganglion of the mouse projects axons to glomeruli in the olfactory bulb. *Eur J Neurosci* **22**, 2649–2654 (2005).

[172] Fleischer, J., Hass, N., Schwarzenbacher, K., Besser, S. & Breer, H. A novel population of neuronal cells expressing the olfactory marker protein (omp) in the anterior/dorsal region of the nasal cavity. *Histochem Cell Biol* **125**, 337–349 (2006).

[173] Brechbuhl, J., Klaey, M. & Broillet, M.-C. Grueneberg ganglion cells mediate alarm pheromone detection in mice. *Science* **321**, 1092–1095 (2008).

[174] Liu, C. Y., Fraser, S. E. & Koos, D. S. Grueneberg ganglion olfactory subsystem employs a cgmp signaling pathway. *J Comp Neurol* **516**, 36–48 (2009).

[175] Fleischer, J., Schwarzenbacher, K., Besser, S., Hass, N. & Breer, H. Olfactory receptors and signalling elements in the grueneberg ganglion. *J Neurochem* **98**, 543–554 (2006).

[176] Fleischer, J., Schwarzenbacher, K. & Breer, H. Expression of trace amine–associated receptors in the grueneberg ganglion. *Chemical Senses* **32**, 623–631 (2007).

[177] Schmid, A., Pyrski, M., Biel, M., Leinders-Zufall, T. & Zufall, F. Grueneberg ganglion neurons are finely tuned cold sensors. *J Neurosci* **30**, 7563–7568 (2010).

[178] Chao, Y.-C. *et al.* Receptor guanylyl cyclase-g is a novel thermosensory protein activated by cool temperatures. *EMBO J* **34**, 294–306 (2015).

[179] Serizawa, S., Miyamichi, K. & Sakano, H. One neuron-one receptor rule in the mouse olfactory system. *Trends in Genetics* **20**, 648 – 653 (2004).

[180] Eggan, K. *et al.* Mice cloned from olfactory sensory neurons. *Nature* **428**, 44–49 (2004).

[181] Li, J., Ishii, T., Feinstein, P. & Mombaerts, P. Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature* **428**, 393–399 (2004).

[182] Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* **302**, 2088–2094 (2003).

[183] Fuss, S. H., Omura, M. & Mombaerts, P. Local and cis effects of the h element on expression of odorant receptor genes in mouse. *Cell* **130**, 373–384 (2007).

[184] Nishizumi, H., Kumasaka, K., Inoue, N., Nakashima, A. & Sakano, H. 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**, 20067–20072 (2007).

[185] Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).

[186] Khan, M., Vaes, E. & Mombaerts, P. Regulation of the probability of mouse odorant receptor gene choice. *Cell* **147**, 907–921 (2011).

[187] Markenscoff-Papadimitriou, E. *et al.* Enhancer interaction networks as a means for singular olfactory receptor expression. *Cell* **159**, 543–557 (2014).

[188] Vassalli, A., Feinstein, P. & Mombaerts, P. Homeodomain binding motifs modulate the probability of odorant receptor gene choice in transgenic mice. *Mol Cell Neurosci* **46**, 381–396 (2011).

[189] Zhang, Y.-Q., Breer, H. & Strotmann, J. Promotor elements governing the clustered expression pattern of odorant receptor genes. *Mol Cell Neurosci* **36**, 95–107 (2007).

[190] Rothman, A., Feinstein, P., Hirota, J. & Mombaerts, P. The promoter of the mouse odorant receptor gene m71. *Mol Cell Neurosci* **28**, 535–546 (2005).

[191] Hoppe, R., Breer, H. & Strotmann, J. Promoter motifs of olfactory receptor genes expressed in distinct topographic patterns. *Genomics* **87**, 711 – 723 (2006).

[192] Michaloski, J. S., Galante, P. A. F. & Malnic, B. Identification of potential regulatory motifs in odorant receptor genes by analysis of promoter sequences. *Genome Res* **16**, 1091–1098 (2006).

[193] Plessy, C. *et al.* Promoter architecture of mouse olfactory receptor genes. *Genome Res* **22**, 486–497 (2012).

[194] Clowney, E. J. *et al.* High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Res* **21**, 1249–1259 (2011).

[195] Young, J. M., Luche, R. M. & Trask, B. J. 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**, 561 (2011).

[196] Hirota, J. & Mombaerts, P. The lim-homeodomain protein lhx2 is required for complete development of mouse olfactory sensory neurons. *Proc Natl Acad Sci U S A* **101**, 8751–8755 (2004).

[197] Kolterud, A., Alenius, M., Carlsson, L. & Bohm, S. The lim homeobox gene lhx2 is required for olfactory sensory neuron identity. *Development* **131**, 5319–5326 (2004).

[198] Hirota, J., Omura, M. & Mombaerts, P. Differential impact of lhx2 deficiency on expression of class i and class ii odorant receptor genes in mouse. *Mol Cell Neurosci* **34**, 679–688 (2007).

[199] McIntyre, J. C., Bose, S. C., Stromberg, A. J. & McClintock, T. S. Emx2 stimulates odorant receptor gene expression. *Chem Senses* **33**, 825–837 (2008).

[200] Nguyen, M. Q., Zhou, Z., Marks, C. A., Ryba, N. J. P. & Belluscio, L. Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell* **131**, 1009–1017 (2007).

[201] Fleischmann, A., Abdus-Saboor, I., Sayed, A. & Shykind, B. Functional interrogation of an odorant receptor locus reveals multiple axes of transcriptional regulation. *PLoS Biol* **11**, e1001568 (2013).

[202] McClintock, T. S. Achieving singularity in mammalian odorant receptor gene choice. *Chemical Senses* **35**, 447–457 (2010).

[203] Magklara, A. *et al.* An epigenetic signature for monoallelic olfactory receptor expression. *Cell* **145**, 555–570 (2011).

[204] Clowney, E. J. *et al.* Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell* **151**, 724–737 (2012).

[205] Lyons, D. B. *et al.* An epigenetic trap stabilizes singular olfactory receptor expression. *Cell* **154**, 325–336 (2013).

[206] Lewcock, J. W. & Reed, R. R. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc Natl Acad Sci U S A* **101**, 1069–1074 (2004).

[207] Shykind, B. M. *et al.* Gene switching and the stability of odorant receptor gene choice. *Cell* **117**, 801 – 815 (2004).

[208] Strotmann, J., Bader, A., Luche, H., Fehling, H. J. & Breer, H. The patch-like pattern of or37 receptors is formed by turning off gene expression in non-appropriate areas. *Molecular and Cellular Neuroscience* **41**, 474 – 485 (2009).

[209] Dalton, R. P., Lyons, D. B. & Lomvardas, S. Co-opting the unfolded protein response to elicit olfactory receptor feedback. *Cell* **155**, 321–332 (2013).

[210] Rodriguez, I. Singular expression of olfactory receptor genes. *Cell* **155**, 274 – 277 (2013).

[211] Tan, L., Zong, C. & Xie, X. S. Rare event of histone demethylation can initiate singular gene expression of olfactory receptors. *Proceedings of the National Academy of Sciences* **110**, 21148–21152 (2013).

[212] Zhao, H. *et al.* Functional expression of a mammalian odorant receptor. *Science* **279**, 237–242 (1998).

[213] Sato, T., Hirono, J., Tonoike, M. & Takebayashi, M. Tuning specificities to aliphatic odorants in mouse olfactory receptor neurons and their local distribution. *J Neurophysiol* **72**, 2980–2989 (1994).

[214] Duchamp-Viret, P., Chaput, M. A. & Duchamp, A. Odor response properties of rat olfactory receptor neurons. *Science* **284**, 2171–2174 (1999).

[215] Kajiya, K. *et al.* Molecular bases of odor discrimination: Reconstitution of olfactory receptors that recognize overlapping sets of odorants. *The Journal of Neuroscience* **21**, 6018–6025 (2001).

[216] Bozza, T., Feinstein, P., Zheng, C. & Mombaerts, P. Odorant receptor expression defines functional units in the mouse olfactory system. *J Neurosci* **22**, 3033–3043 (2002).

[217] Rubin, B. D. & Katz, L. C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1999).

[218] Ma, M. & Shepherd, G. M. Functional mosaic organization of mouse olfactory receptor neurons. *Proc Natl Acad Sci U S A* **97**, 12869–12874 (2000).

[219] Belluscio, L., Lodovichi, C., Feinstein, P., Mombaerts, P. & Katz, L. C. Odorant receptors instruct functional circuitry in the mouse olfactory bulb. *Nature* **419**, 296–300 (2002).

[220] Araneda, R. C., Kini, A. D. & Firestein, S. The molecular receptive range of an odorant receptor. *Nat Neurosci* **3**, 1248–1255 (2000).

[221] Katada, S., Hirokawa, T., Oka, Y., Suwa, M. & Touhara, K. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: Mapping the odorant-binding site. *The Journal of Neuroscience* **25**, 1806–1815 (2005).

[222] Araneda, R. C., Peterlin, Z., Zhang, X., Chesler, A. & Firestein, S. A pharmacological profile of the aldehyde receptor repertoire in rat olfactory epithelium. *The Journal of Physiology* **555**, 743–756 (2004).

[223] Nara, K., Saraiva, L. R., Ye, X. & Buck, L. B. A large-scale analysis of odor coding in the olfactory epithelium. *J Neurosci* **31**, 9179–9191 (2011).

[224] Krautwurst, D., Yau, K. W. & Reed, R. R. Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell* **95**, 917–926 (1998).

[225] Rubin, B. D. & Katz, L. C. Spatial coding of enantiomers in the rat olfactory bulb. *Nat Neurosci* **4**, 355–356 (2001).

[226] Hamana, H., Hirono, J., Kizumi, M. & Sato, T. Sensitivity-dependent hierarchical receptor codes for odors. *Chemical Senses* **28**, 87–104 (2003).

[227] Touhara, K. *et al.* Functional identification and reconstitution of an odorant receptor in single olfactory neurons. *Proceedings of the National Academy of Sciences* **96**, 4040–4045 (1999).

[228] Oka, Y. *et al.* Odorant receptor map in the mouse olfactory bulb: in vivo sensitivity and specificity of receptor-defined glomeruli. *Neuron* **52**, 857–869 (2006).

[229] Tsuboi, A. *et al.* Two highly homologous mouse odorant receptors encoded by tandemly-linked mor29a and mor29b genes respond differently to phenyl ethers. *Eur J Neurosci* **33**, 205–213 (2011).

[230] Peterlin, Z., Firestein, S. & Rogers, M. E. The state of the art of odorant receptor deorphanization: a report from the orphanage. *J Gen Physiol* **143**, 527–542 (2014).

[231] Shirasu, M. *et al.* Olfactory receptor and neural pathway responsible for highly selective sensing of musk odors. *Neuron* **81**, 165–178 (2014).

[232] McClintock, T. S. *et al.* In vivo identification of eugenol-responsive and muscone-responsive mouse odorant receptors. *J Neurosci* **34**, 15669–15678 (2014).

[233] Gaillard, I. *et al.* A single olfactory receptor specifically binds a set of odorant molecules. *Eur J Neurosci* **15**, 409–418 (2002).

[234] Lu, M., Echeverri, F. & Moyer, B. D. Endoplasmic reticulum retention, degradation, and aggregation of olfactory g-protein coupled receptors. *Traffic* **4**, 416–433 (2003).

[235] Saito, H., Kubota, M., Roberts, R. W., Chi, Q. & Matsunami, H. Rtp family members induce functional expression of mammalian odorant receptors. *Cell* **119**, 679 – 691 (2004).

[236] Zhuang, H. & Matsunami, H. Synergism of accessory factors in functional expression of mammalian odorant receptors. *Journal of Biological Chemistry* **282**, 15284–15293 (2007).

[237] Von Dannecker, L. E. C., Mercadante, A. F. & Malnic, B. Ric-8b, an olfactory putative gtp exchange factor, amplifies signal transduction through the olfactory-specific g-protein galpha-olf. *The Journal of Neuroscience* **25**, 3793–3800 (2005).

[238] Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. D. Odor coding by a mammalian receptor repertoire. *Sci Signal* **2**, ra9 (2009).

[239] Shepard, B. D., Natarajan, N., Protzko, R. J., Acres, O. W. & Pluznick, J. L. A cleavable n-terminal signal peptide promotes widespread olfactory receptor surface expression in hek293t cells. *PLoS One* **8**, e68758 (2013).

[240] Shirokova, E. *et al.* Identification of specific ligands for orphan olfactory receptors. g protein-dependent agonism and antagonism of odorants. *J Biol Chem* **280**, 11807–11815 (2005).

[241] Mainland, J. D. *et al.* The missense of smell: functional variability in the human odorant receptor repertoire. *Nat Neurosci* **17**, 114–120 (2014).

[242] Oka, Y., Omura, M., Kataoka, H. & Touhara, K. Olfactory receptor antagonism between odorants. *The EMBO journal* **23**, 120–126 (2004).

[243] Oka, Y., Nakamura, A., Watanabe, H. & Touhara, K. An odorant derivative as an antagonist for an olfactory receptor. *Chemical Senses* **29**, 815–822 (2004).

[244] Peterlin, Z. *et al.* The importance of odorant conformation to the binding and activation of a representative olfactory receptor. *Chemistry /& biology* **15**, 1317–1327 (2008).

[245] Zufall, F. & Leinders-Zufall, T. The cellular and molecular basis of odor adaptation. *Chemical Senses* **25**, 473–481 (2000).

[246] Bradley, J., Bonigk, W., Yau, K.-W. & Frings, S. Calmodulin permanently associates with rat olfactory cng channels under native conditions. *Nat Neurosci* **7**, 705–710 (2004).

[247] Bradley, J., Reuter, D. & Frings, S. Facilitation of calmodulin-mediated odor adaptation by camp-gated channel subunits. *Science* **294**, 2176–2178 (2001).

[248] Song, Y. *et al.* Olfactory cng channel desensitization by ca2+/cam via the b1b subunit affects response termination but not sensitivity to recurring stimulation. *Neuron* **58**, 374–386 (2008).

[249] Yan, C. *et al.* Molecular cloning and characterization of a calmodulin-dependent phosphodiesterase enriched in olfactory sensory neurons. *Proc Natl Acad Sci U S A* **92**, 9677–9681 (1995).

[250] Wei, J. *et al.* Phosphorylation and inhibition of olfactory adenylyl cyclase by cam kinase ii in neurons: a mechanism for attenuation of olfactory signals. *Neuron* **21**, 495–504 (1998).

[251] Leinders-Zufall, T., Ma, M. & Zufall, F. Impaired odor adaptation in olfactory receptor neurons after inhibition of ca2+/calmodulin kinase ii. *J Neurosci* **19**, RC19 (1999).

[252] Dawson, T. M. *et al.* Beta-adrenergic receptor kinase-2 and beta-arrestin-2 as mediators of odorant-induced desensitization. *Science* **259**, 825–829 (1993).

[253] Mashukova, A., Spehr, M., Hatt, H. & Neuhaus, E. M. Beta-arrestin2-mediated internalization of mammalian odorant receptors. *J Neurosci* **26**, 9902–9912 (2006).

[254] Boekhoff, I. *et al.* Olfactory desensitization requires membrane targeting of receptor kinase mediated by beta gamma-subunits of heterotrimeric g proteins. *J Biol Chem* **269**, 37–40 (1994).

[255] Nozawa, M., Kawahara, Y. & Nei, M. Genomic drift and copy number variation of sensory receptor genes in humans. *Proceedings of the National Academy of Sciences* **104**, 20421–20426 (2007).

[256] Hasin, Y. *et al.* High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genet* **4**, e1000249 (2008).

[257] Waszak, S. M. *et al.* Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory receptor gene content diversity. *PLoS Comput Biol* **6**, e1000988 (2010).

[258] Menashe, I., Man, O., Lancet, D. & Gilad, Y. Different noses for different people. *Nat Genet* **34**, 143–144 (2003).

[259] Gilad, Y. & Lancet, D. Population differences in the human functional olfactory repertoire. *Mol Biol Evol* **20**, 307–314 (2003).

[260] Olender, T. *et al.* Personal receptor repertoires: olfaction as a model. *BMC Genomics* **13**, 414 (2012).

[261] Menashe, I. *et al.* Genetic elucidation of human hyperosmia to isovaleric acid. *PLoS Biol* **5**, e284 (2007).

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

[263] Griff, I. C. & Reed, R. R. The genetic basis for specific anosmia to isovaleric acid in the mouse. *Cell* **83**, 407–414 (1995).

[264] Adipietro, K. A., Mainland, J. D. & Matsunami, H. Functional evolution of mammalian odorant receptors. *PLoS Genet* **8**, e1002821 (2012).

[265] Whissell-Buechy, D. & Amoore, J. E. Odour-blindness to musk: simple recessive inheritance. *Nature* **245**, 157–158 (1973).

[266] Jaeger, S. R. *et al.* A mendelian trait for olfactory sensitivity affects odor experience and food selection. *Current Biology* **23**, 1601 – 1605 (2013).

[267] McRae, J. F. *et al.* Genetic variation in the odorant receptor or2j3 is associated with the ability to detect the "grassy" smelling odor, cis-3-hexen-1-ol. *Chem Senses* **37**, 585–593 (2012).

[268] Lunde, K. *et al.* Genetic variation of an odorant receptor or7d4 and sensory perception of cooked meat containing androstenone. *PLoS One* **7**, e35259 (2012).

[269] Knaapila, A. *et al.* A genome-wide study on the perception of the odorants androstenone and galaxolide. *Chem Senses* **37**, 541–552 (2012).

[270] Keller, A., Zhuang, H., Chi, Q., Vosshall, L. B. & Matsunami, H. Genetic variation in a human odorant receptor alters odour perception. *Nature* **449**, 468–472 (2007).

[271] Logan, D. W. Do you smell what i smell? genetic variation in olfactory perception. *Biochem Soc Trans* **42**, 861–865 (2014).

[272] Frumin, I., Sobel, N. & Gilad, Y. Does a unique olfactory genome imply a unique olfactory world? *Nat Neurosci* **17**, 6–8 (2014).

[273] Lledo, P.-M., Alonso, M. & Grubb, M. S. Adult neurogenesis and functional plasticity in neuronal circuits. *Nat Rev Neurosci* **7**, 179–193 (2006).

[274] Yamaguchi, M. & Mori, K. Critical period for sensory experience-dependent survival of newly generated granule cells in the adult mouse olfactory bulb. *Proc Natl Acad Sci U S A* **102**, 9697–9702 (2005).

[275] Rochefort, C., Gheusi, G., Vincent, J.-D. & Lledo, P.-M. Enriched odor exposure increases the number of newborn neurons in the adult olfactory bulb and improves odor memory. *J Neurosci* **22**, 2679–2689 (2002).

[276] Rochefort, C. & Lledo, P.-M. Short-term survival of newborn neurons in the adult olfactory bulb after exposure to a complex odor environment. *Eur J Neurosci* **22**, 2863–2870 (2005).

[277] Alonso, M. *et al.* Olfactory discrimination learning increases the survival of adult-born neurons in the olfactory bulb. *J Neurosci* **26**, 10508–10513 (2006).

[278] Woo, C. C., Hingco, E. E., Taylor, G. E. & Leon, M. Exposure to a broad range of odorants decreases cell mortality in the olfactory bulb. *Neuroreport* **17**, 817–821 (2006).

[279] Magavi, S. S. P., Mitchell, B. D., Szentirmai, O., Carter, B. S. & Macklis, J. D. Adult-born and preexisting olfactory granule neurons undergo distinct experience-dependent modifications of their olfactory responses in vivo. *J Neurosci* **25**, 10729–10739 (2005).

[280] Sultan, S. *et al.* Learning-dependent neurogenesis in the olfactory bulb determines long-term olfactory memory. *FASEB J* **24**, 2355–2363 (2010).

[281] Sultan, S., Rey, N., Sacquet, J., Mandairon, N. & Didier, A. Newborn neurons in the olfactory bulb selected for long-term survival through olfactory learning are prematurely suppressed when the olfactory memory is erased. *J Neurosci* **31**, 14893–14898 (2011).

[282] Mandairon, N., Stack, C., Kiselycznyk, C. & Linster, C. Enrichment to odors improves olfactory discrimination in adult rats. *Behav Neurosci* **120**, 173–179 (2006).

[283] Escanilla, O., Mandairon, N. & Linster, C. Odor-reward learning and enrichment have similar effects on odor perception. *Physiol Behav* **94**, 621–626 (2008).

[284] Veyrac, A. *et al.* Novelty determines the effects of olfactory enrichment on memory and neurogenesis through noradrenergic mechanisms. *Neuropsychopharmacology* **34**, 786–795 (2009).

[285] Gusmao, I. D. *et al.* Odor-enriched environment rescues long-term social memory, but does not improve olfaction in social isolated adult mice. *Behav Brain Res* **228**, 440–446 (2012).

[286] Watt, W. C. *et al.* Odorant stimulation enhances survival of olfactory sensory neurons via mapk and creb. *Neuron* **41**, 955–967 (2004).

[287] Kim, S. Y., Yoo, S.-J., Ronnett, G. V., Kim, E.-K. & Moon, C. Odorant stimulation promotes survival of rodent olfactory receptor neurons via pi3k/akt activation and bcl-2 expression. *Mol Cells* **38**, 535–539 (2015).

[288] François, A. *et al.* Early survival factor deprivation in the olfactory epithelium enhances activity-dependent survival. *Frontiers in Cellular Neuroscience* **7** (2013).

[289] Zhao, H. & Reed, R. R. X inactivation of the ocnc1 channel gene reveals a role for activity-dependent competition in the olfactory system. *Cell* **104**, 651–660 (2001).

[290] Zhao, S. *et al.* Activity-dependent modulation of odorant receptor gene expression in the mouse olfactory epithelium. *PLoS One* **8**, e69862 (2013).

[291] Coppola, D. M. & Waggener, C. T. The effects of unilateral naris occlusion on gene expression profiles in mouse olfactory mucosa. *J Mol Neurosci* **47**, 604–618 (2012).

[292] Santoro, S. W. & Dulac, C. The activity-dependent histone variant h2be modulates the life span of olfactory neurons. *Elife* **1**, e00070 (2012). NLM: Original DateCompleted: 20121217.

[293] Cavallin, M. A., Powell, K., Biju, K. C. & Fadool, D. A. State-dependent sculpting of olfactory sensory neurons is attributed to sensory enrichment, odor deprivation, and aging. *Neurosci Lett* **483**, 90–95 (2010).

[294] Cadiou, H. *et al.* Postnatal odorant exposure induces peripheral olfactory plasticity at the cellular level. *J Neurosci* **34**, 4857–4870 (2014).

[295] Ibarra-Soria, X., Levitin, M. O., Saraiva, L. R. & Logan, D. W. The olfactory transcriptomes of mice. *PLoS Genet* **10**, e1004593 (2014).

[296] Niimura, Y., Matsui, A. & Touhara, K. Extreme expansion of the olfactory receptor gene repertoire in african elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res* **24**, 1485–1496 (2014).

[297] Spehr, M. *et al.* Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science* **299**, 2054–2058 (2003).

[298] Fukuda, N., Yomogida, K., Okabe, M. & Touhara, K. Functional characterization of a mouse testicular olfactory receptor and its role in chemosensing and in regulation of sperm motility. *J Cell Sci* **117**, 5835–5845 (2004).

[299] Zhang, X. *et al.* High-throughput microarray detection of olfactory receptor gene expression in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14168–14173 (2004).

[300] Zhang, X., Marcucci, F. & Firestein, S. High-throughput microarray detection of vomeronasal receptor gene expression in rodents. *Frontiers in Neuroscience* **4** (2010).

[301] Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. Rna-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome research* **18**, 1509–1517 (2008).

[302] Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **26**, 317–325 (2008).

[303] Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nat Meth* **5**, 621–628 (2008).

[304] Khan, M., Vaes, E. & Mombaerts, P. Temporal patterns of odorant receptor gene expression in adult and aged mice. *Mol Cell Neurosci* **57**, 120–129 (2013).

[305] Hebenstreit, D. *et al.* Rna sequencing reveals two major classes of gene expression levels in metazoan cells. *Molecular systems biology* **7**, 497 (2011).

[306] Xie, S. Y., Feinstein, P. & Mombaerts, P. Characterization of a cluster comprising approximately 100 odorant receptor genes in mouse. *Mamm Genome* **11**, 1070–1078 (2000).

[307] Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using rna-seq. *Bioinformatics* **27**, 2325–2329 (2011).

[308] Saraiva, L. R. *et al.* Hierarchical deconstruction of mouse olfactory sensory neurons: from whole mucosa to single-cell rna-seq. *Sci Rep* **5**, 18178 (2015).

[309] Potter, S. M. *et al.* Structure and emergence of specific olfactory glomeruli in the mouse. *J Neurosci* **21**, 9713–9723 (2001).

[310] Keller, A. & Margolis, F. L. Immunological studies of the rat olfactory marker protein. *J Neurochem* **24**, 1101–1106 (1975).

[311] Monti-Graziadei, G. A., Margolis, F. L., Harding, J. W. & Graziadei, P. P. Immunocytochemistry of the olfactory marker protein. *J Histochem Cytochem* **25**, 1311–1316 (1977).

[312] Rodriguez-Gil, D. J. *et al.* Odorant receptors regulate the final glomerular coalescence of olfactory sensory neuron axons. *Proceedings of the National Academy of Sciences* **112**, 5821–5826 (2015).

[313] Buiakova, O. I. *et al.* Olfactory marker protein (omp) gene deletion causes altered physiological activity of olfactory sensory neurons. *Proc Natl Acad Sci U S A* **93**, 9858–9863 (1996).

[314] Sammeta, N., Yu, T.-T., Bose, S. C. & McClintock, T. S. Mouse olfactory sensory neurons express 10,000 genes. *J Comp Neurol* **502**, 1138–1156 (2007).

[315] Nickell, M. D., Breheny, P., Stromberg, A. J. & McClintock, T. S. Genomics of mature and immature olfactory sensory neurons. *J Comp Neurol* **520**, 2608–2629 (2012).

[316] Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* **16**, 133–145 (2015).

[317] Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

[318] Brennecke, P. *et al.* Accounting for technical noise in single-cell rna-seq experiments. *Nat Methods* **10**, 1093–1095 (2013).

[319] Mahata, B. *et al.* Single-cell rna sequencing reveals t helper cells synthesizing steroids de novo to contribute to immune homeostasis. *Cell Rep* **7**, 1130–1142 (2014).

[320] Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell rna-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* **33**, 155–160 (2015).

[321] Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).

[322] Omura, M. & Mombaerts, P. Trpc2-expressing sensory neurons in the mouse main olfactory epithelium of type b express the soluble guanylate cyclase gucy1b2. *Molecular and cellular neurosciences* **65**, 114–124 (2015).

[323] Ferrero, D. M. *et al.* A juvenile mouse pheromone inhibits sexual behaviour through the vomeronasal system. *Nature* **502**, 368–371 (2013).

[324] Stowers, L. & Logan, D. W. Sexual dimorphism in olfactory signaling. *Curr Opin Neurobiol* **20**, 770–775 (2010).

[325] Bressel, O. C., Khan, M. & Mombaerts, P. 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* (2015).

[326] Kwak, J. *et al.* Differential binding between volatile ligands and major urinary proteins due to genetic variation in mice. *Physiol Behav* **107**, 112–120 (2012).

[327] Yamaguchi, M. *et al.* Distinctive urinary odors governed by the major histocompatibility locus of the mouse. *Proc Natl Acad Sci U S A* **78**, 5817–5820 (1981).

[328] Goncalves, A. *et al.* Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res* **22**, 2376–2384 (2012).

[329] Shiao, M.-S. *et al.* Transcriptomes of mouse olfactory epithelium reveal sexual differences in odorant detection. *Genome Biol Evol* **4**, 703–712 (2012).

[330] Kanageswaran, N. *et al.* Deep sequencing of the murine olfactory receptor neuron transcriptome. *PLoS One* **10**, e0113170 (2015).

[331] Shum, E. Y., Espinoza, J. L., Ramaiah, M. & Wilkinson, M. F. Identification of novel post-transcriptional features in olfactory receptor family mrnas. *Nucleic Acids Res* (2015).

[332] Kuhlmann, K. *et al.* The membrane proteome of sensory cilia to the depth of olfactory receptors. *Mol Cell Proteomics* **13**, 1828–1843 (2014).

[333] Kang, N. & Koo, J. Olfactory receptors in non-chemosensory tissues. *BMB Rep* **45**, 612–622 (2012).

[334] Griffin, C. A., Kafadar, K. A. & Pavlath, G. K. Mor23 promotes muscle regeneration and regulates cell adhesion and migration. *Dev Cell* **17**, 649–661 (2009).

[335] Braun, T., Voland, P., Kunz, L., Prinz, C. & Gratzl, M. Enterochromaffin cells of the human gut: sensors for spices and odorants. *Gastroenterology* **132**, 1890–1901 (2007).

[336] Wang, F., Kessels, H. W. & Hu, H. The mouse that roared: neural mechanisms of social hierarchy. *Trends Neurosci* **37**, 674–682 (2014).

[337] Dey, S. *et al.* Cyclic regulation of sensory perception by a female hormone alters behavior. *Cell* **161**, 1334–1344 (2015).

[338] Yang, H., Bell, T. A., Churchill, G. A. & Pardo-Manuel de Villena, F. On the subspecific origin of the laboratory mouse. *Nat Genet* **39**, 1100–1107 (2007).

[339] Jorde, L. B. & Wooding, S. P. Genetic variation, classification and 'race'. *Nat Genet* **36**, S28–33 (2004).

[340] Pezer, Z., Harr, B., Teschke, M., Babiker, H. & Tautz, D. Divergence patterns of genic copy number variation in natural populations of the house mouse (mus musculus domesticus) reveal three conserved genes with major population-specific expansions. *Genome Res* **25**, 1114–1124 (2015).

[341] Liu, K. J. *et al.* Interspecific introgressive origin of genomic diversity in the house mouse. *Proc Natl Acad Sci U S A* **112**, 196–201 (2015).

[342] Secundo, L. *et al.* Individual olfactory perception reveals meaningful nonolfactory genetic information. *Proc Natl Acad Sci U S A* **112**, 8750–8755 (2015).

[343] Fleischmann, A. *et al.* Mice with a "monoclonal nose": perturbations in an olfactory map impair odor discrimination. *Neuron* **60**, 1068–1081 (2008).

[344] von der Weid, B. *et al.* Large-scale transcriptional profiling of chemosensory neurons identifies receptor-ligand pairs in vivo. *Nat Neurosci* (2015).

[345] Jiang, Y. *et al.* Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo. *Nat Neurosci* (2015).

[346] Li, H. *et al.* The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–2079 (2009).

[347] Dobin, A. *et al.* Star: ultrafast universal rna-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

[348] Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).

[349] Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (igv): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192 (2013).

[350] Anders, S., Pyl, P. T. & Huber, W. Htseq–a python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

[351] Warnes, G. R. *et al. gplots: Various R Programming Tools for Plotting Data* (2015). R package version 2.17.0.

[352] Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol* **15**, 550 (2014).

[353] Benaglia, T., Chauveau, D., Hunter, D. R. & Young, D. mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software* **32**, 1–29 (2009).

[354] Backes, C. *et al.* Genetrail–advanced gene set enrichment analysis. *Nucleic Acids Res* **35**, W186–92 (2007).

[355] Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).

[356] McLaren, W. *et al.* Deriving the consequences of genomic variants with the ensembl api and snp effect predictor. *Bioinformatics* **26**, 2069–2070 (2010).

[357] Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol* **10**, R25 (2009).

[358] Quinlan, A. R. & Hall, I. M. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

[359] Munger, S. C. *et al.* Rna-seq alignment to individualized genomes improves transcript abundance estimates in multiparent populations. *Genetics* **198**, 59–73 (2014).

[360] Micallef, L. & Rodgers, P. eulerape: drawing area-proportional 3-venn diagrams using ellipses. *PLoS One* **9**, e101717 (2014).

Appendix A

Methods

Sample collection and RNA extraction.

All mice used were group housed. The details of the strain, age and sex of each sample can be found in Table B.1 in Appendix B. In the case of the VNO samples, each biological replicate was the pool of three animals. All WOM samples were obtained from a single animal, except the pup WOM samples, which were the pool of 3 or 4 individuals. Tissue was dissected and immediately homogenised in lysis RLT buffer (Qiagen) using a disposable RNAse free plastic grinder, except for the pup samples, which were stored in RNAlater. Total RNA was extracted using the RNeasy mini kit (Qiagen) with on-column DNAse digestion, following the manufacturer's protocol. Tissue homogenisation was performed on a QIAshredder column. All RNA was subsequently quantified with a spectrophotometer and visualised for quality by RNA integrity analysis.

Library preparation and sequencing.

mRNA was prepared for sequencing using the TruSeq RNA sample preparation kit (Illumina). All RNA sequencing was paired-end. The details of the specific Illumina platform used, read length and data strandedness are in Table B.1 in Appendix B. All raw sequencing data are available through the European Nucleotide Archive (ENA); the corresponding accession numbers for each sample can be found in Tables B.1 and B.4 in Appendix B.

RNAseq data processing and mapping.

BAM files were processed using SAMtools[346] and Picard tools version 1.64

(http://broadinstitute.github.io/picard).

Sequencing data were aligned with STAR 2.3[347]. Prior to mapping, the genome index was built with the GTF annotation file under *–sjdbGTFfile* and with option *–sjdbOverhang* 99. Mapping was performed to the GRCm38 mouse reference genome plus the ERCC spike-in sequences, with options *–outFilterMultimapNmax* 1000 *–outFilterMismatchNmax* 4 *–outFilterMatchNmin* 100 *–alignIntronMax* 50000 *–alignMatesGapMax* 50500 *–outSAMstrandField* intronMotif *–outFilterType* BySJout.

The annotation used for the first dataset presented in Chapter 2 was from the Ensembl mouse genome database, version 68 (http://jul2012.archive.ensembl.org/info/data/ftp/index.html). After reconstruction of full-length gene models for the VR and OR gene repertoires (*see below*), the GTF file from the Ensembl mouse genome database version 72 (http://jun2013.archive.ensembl.org/info/data/ftp/index.html) was modified to include all these reconstructed gene models. Additionally, the set of transcripts reported for *Trpc2* contain both short and long isoforms of the gene; the long isoforms represent a fusion with a different gene and were therefore removed[1]. All data was subsequently mapped and analysed using this annotation file (including the initial dataset which was reanalysed). In the case of the single-cell RNAseq data (Chapter 3), the gene *Gm20715* (a predicted gene that undergoes nonsense mediated decay) was also removed from the GTF file because it overlaps with *Olfr1344*; this overlap causes all the reads aligned to the OR to be deemed ambiguous.

Sequencing data was visualised using the Integrative Genomics Viewer (IGV)[348, 349].

Gene expression level estimation and data analysis.

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*[350]. All multi-mapped fragments were discarded. Data analysis, statistical testing and plotting was carried out in R (http://www.R-project.org). All the heatmaps were produced with the *gplots* package[351] using the $\log_{10}$ transformed normalised counts + 1.

[1]Transcripts removed: ENSMUST00000084843, ENSMUST00000094129, ENSMUST00000094130, ENSMUST00000106950, ENSMUST00000123372, ENSMUST00000125197, ENSMUST00000139104, ENSMUST00000140395, ENSMUST00000141646, ENSMUST00000142629, ENSMUST00000143839, ENSMUST00000146450, ENSMUST00000153176.

RNAseq data normalisation.

Raw counts were normalised to account for sequencing depth between samples, using the procedure implemented in the DESeq2 package[352]. Size factors were calculated with *estimateSizeFactorsForMatrix* and then used to divide the raw counts. For the single-cell data, ERCC spike-ins were not included for data normalisation.

To compare OR expression levels between datasets, normalisation to account for the number of OSNs present in the WOM samples was carried out subsequent to depth normalisation (data presented in Chapters 4 and 5). For this, a method proposed by Khan et al. [304] was used. Five different marker genes were considered, all of which are expressed exclusively in mature OSNs: *Adcy3*, *Ano2*, *Cnga2*, *Gnal* and *Omp*. Further, these have been shown to be expressed at stable levels[304]. To normalise for OSN number the following procedure was applied to the OR normalised counts. First, the correlation between the expression of each of the marker genes and the total number of counts in OR genes was calculated, and all those marker genes with strong correlation values were used. Second, the geometric mean of all marker genes was calculated for each sample. Then, the average of all means was obtained, and divided by each individual mean; this results in the generation of size factors. Third, the OR normalised counts were multiplied by the corresponding size factor.

Differential expression analysis.

To test for differential expression I used DESeq2 1.8.1 with standard parameters. When applied to the single-cell data, the parameter *minReplicatesForReplace* was set to *Inf* to turn off the automatic outlier replacement. Genes were considered differentially expressed if they had an adjusted p-value of 0.05 or less (equivalent to a false discovery rate of 5%). To test for differential expression on the OR repertoire (Chapters 4 and 5) the double normalised counts (accounting for OSN number per sample) were provided directly, and the *normalizationFactors* function was used with size factors of 1 to turn off further normalisation.

Fitting normal distributions to bimodal data.

To deconvolve bimodal distributions into two normal-like distributions I used Gaussian mixture models, through the expectation-maximisation algorithm of the *mixtools* Bioconductor package[353]. In all cases the algorithm converged to optimal values.

Gene enrichment analysis.

To find functional terms enriched in the lists of differentially expressed genes I used GeneTrail with 'Over-/Under-representation Analysis' with default parameters[354]. The background provided were all those genes tested for differential expression (those with an adjusted p-value different to NA).

Microarray profiling.

RNA was extracted from the VNO and WOM of six C57BL/6J males of 10 weeks of age as described above. Profiling was performed on the Illumina MouseWG-6 v2.0 Expression BeadChip following the manufacturer's instructions. Variance stabilising transformation was applied to the data obtained from BeadStudio, which was then quantile normalised using the Bioconductor R package, *lumi*[355].

Recovery of unannotated receptor genes

To recover the entirety of the VR gene repertoire, I took the cDNA sequences as reported[65, 136] and locally aligned them to the mouse genome with BLAST. Then I identified those alignments that overlap genes not annotated as VRs with 100% identity, and changed their name while preserving the Ensembl identifier. In all cases the coordinates obtained from the alignments were concordant with the annotation. A list detailing the gene names that were changed is reported in Table B.3 in Appendix B. Furthermore, 19 additional predicted genes have high identity alignments to other VR sequences. Similarly, I aligned with BLAST all the OR cDNA sequences present in Ensembl v68 and recovered four predicted genes that share high similarity to other ORs. Although these genes are most likely additional members of the VR and OR gene families, proper annotation with novel gene names is required; these were not included as part of the receptor repertoires.

Reconstruction of novel gene models.

To search for novel genes I performed Reference Annotation Based Transcript (RABT) Assembly, using Cufflinks v2.1.1[307] guided by the Ensembl annotation (version 68), with all six replicates of the VNO and WOM data presented in Chapter 2. Assembled transcripts from the different replicates were combined with Cuffmerge. In order to extract the candidates with greatest probability of encoding protein coding genes, I

cross-referenced all predicted loci to the Ensembl databases using the API[356]. *Ad hoc* perl scripts were used to further refine the gene models produced for VR and OR genes, deleting those predictions that fuse adjacent receptor genes or that are antisense to the annotated gene.

Estimation of gene uniqueness.

To calculate the proportion of sequence that is unique in the genome for each receptor gene, I used a perl script to produce all the 32, 76 and 100 nucleotide-long strings that cover the receptor transcripts, either using the Ensembl v68 annotation or the reconstructed gene models by Cufflinks. These were then aligned to the genome with bowtie version 0.12.8[357] and parameters -v 0 -m 1. The unmapped strings were subsequently aligned to the transcriptome, to account for those that span exon-exon junctions. Finally, *ad hoc* perl scripts were used to consolidate the data and count the number of strings that were unique for each gene. The *uniqueness* of a gene was defined as the number of unique strings over the total number of strings for that gene.

Coverage of OR genes.

To obtain the proportion of the OR gene models covered by the mapped sequencing fragments, the BEDtools 2.16.2[358] program coverageBed was used against a BED file containing the merged exonic regions for all isoforms of each OR gene (obtained with mergeBed). The output was then analysed in R to count all positions with at least one mapped fragment to them.

Allelic Discrimination of OR genes.

To determine the allele expressed for each OR in the single-cell data, the Mouse Genomes Project database release 1410 was queried (http://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1410)[321] to obtain all the SNPs for 129P2 that overlap OR gene models. These positions were visualised on IGV and the numbers of fragments containing each nucleotide were extracted.

Creation of pseudo-reference genomes.

To create psuedo-129 and pseudo-CAST genomes, I mined the Mouse Genomes Project data, release v3 (ftp://ftp-mouse.sanger.ac.uk/REL-1303-SNPs_Indels-GRCm38/) to

obtain all the high-quality SNPs and short indels for the 129S5SvEvBrd and CAST/EiJ strains, respectively. These were imputed into the GRCm38 mouse reference genome using Seqnature[359].

Proportional Venn diagrams.

Venn diagrams with areas proportional to the number of elements represented were created using the eulerAPE version 3 software[360].

Dissecting genetic from environmental effects experiment.

To dissect the influence of the genetic background from the olfactory environment between B6 and 129 animals, C57BL/6N and 129S5 4 to 8-cell stage embryos were transferred into F1 (C57BL/6J×CBA) pseudo-pregnant females, and allowed to develop in this equivalent *in utero* environment. One day after birth, the C57BL/6N and 129S5 litters were cross-fostered to C57BL/6N and 129S5 wild-type mothers, respectively. For this, the mothers were removed from their home cage, and the pups to be cross-fostered were introduced to the home-cage of the foster mother; each pup was gently rubbed with nesting material to transfer some of the odours. Then, the mother was introduced into the cage with the new litter, and observed for at least half an hour to ensure it did not reject the pups; those that did were separated from the litter. Then, a single pup from the other strain was transferred to the cross-fostered litter (the *alien*). At weaning, animals from the same sex as the alien animal were kept, always in a 4:1 ratio between strains. If not enough animals of the correct sex were available in the litter, surplus animals from other litters were used. At 10 weeks of age, the WOM was collected from the alien and a randomly selected cage-mate, and RNA was extracted as described previously.

The details on the strain of the alien and cage-mate for each sequenced sample are as follows:

sample	sex	alien	cage-mate
1	female	B6	129
2	female	129	B6
3	female	129	B6
4	female	B6	129
5	male	129	B6
6	male	B6	129

To test for the effect of the environment on gene expression, I used a likelihood ratio test with DESeq2, to test the model genetics+environment+genetics:environment versus accounting only for the genetics; this revealed two significant genes, both ORs. If, instead, the data from each strain was tested separately for the effect of the environment, one of the genes previously identified was again recovered for the B6 data, and a new gene (ENSMUSG00000063779) was significant for the 129 data.

Allelic discrimination of the F1 RNAseq data.

RNAseq data was processed as described above. Total expression estimates were obtained by mapping the RNAseq data to the B6 or pseudo-CAST genomes, with standard parameters. The expression estimates obtained with each genome were very highly correlated. For the OR repertoire, nearly all the genes (96.23%) differed in less than 10 counts and were almost perfectly correlated (rho = 0.9991006, p-value < 2.2e-16). Thus, by allowing 4 mismatches per paired-end fragment, nearly all reads were able to be mapped regardless of the reference used. Therefore, the data mapped to the B6 reference was used in downstream analyses.

To obtain allele-specific expression estimates, the RNAseq data was mapped to both the B6 and the pseudo-CAST genomes, without mismatches. In this way, those reads that span SNPs, could only map to the genome corresponding to the allele they come from. Subsequent analyses were performed on the OR repertoire only. All reads mapped across each SNP were retrieved with SAMtools[346]. In cases where different transcripts exist, and one of them splices across the SNP, SAMtools reports both the reads that map and splice across the SNP. *Ad hoc* perl scripts were used to retain only reads that contained the SNP (using the cigar string) and that were uniquely mapped. Finally, the number of different reads mapping across all SNPs of each gene was obtained. The results using the data mapped to either the B6 or CAST genomes then provide the number of reads that are specific for each allele.

To normalise for depth of sequencing, the total expression raw data was combined with the estimates from the parental strains, and normalised all together. The OR data was then further normalised to account for the number of OSNs, as described above. The same size factors were used to normalise the expression estimates from SNP positions.

To deconvolve the total expression into allele-specific expression, a ratio of the expression of each allele was obtained from the counts in SNP positions with:

$$\frac{\text{counts in B6}}{\text{counts in B6 + counts in CAST}}$$

Then, the total expression normalised counts were multiplied by the ratio to obtain the B6 expression, and to the inverse of the ratio for the CAST-specific expression. Finally, since those genes with very low number of SNPs and/or very low expression have very few reads spanning SNPs, the information is very limited and the estimated ratio is not robust. Thus, only those genes with normalised counts in SNP positions above the lowest quartile were used (82.5%).

Odour-exposure experiments.

To test the effects of enriching the environment with specific odorants, I selected heptanal, (R)-carvone, eugenol and acetophenone because they all have been shown to activate at least one specific OR gene. All odorants were from Sigma, except for acetophenone which was from Alfa Aesar. The mixture of all four consisted of equimolar proportions of each, diluted in mineral oil (Sigma) for a final concentration of 1mM each.

For the chronic exposure experiments, a couple drops of the odour mixture, or mineral oil only, were applied to a cotton ball with a plastic pasteur pipette, for the *exposed* and *control* groups respectively; 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 odour mix was changed twice a week for a freshly prepared stock. 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.

For the acute exposure experiments, the odour mix or mineral oil was added to the water bottles of the animals. Water bottles were replaced twice a week with freshly prepared ones. 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.

The number of animals analysed in each group were as follows:

CHRONIC						
time-point	**control**		**exposed**		**total**	
	males	**females**	**males**	**females**	**control**	**exposed**
4	4	0	5	0	4	5
10	3	0	4	0	3	4
24	5	5	4	5	10	9

ACUTE						
time-point	**control**		**exposed**		**total**	
	males	**females**	**males**	**females**	**control**	**exposed**
1	8		8		8	8
4	5	3	5	5	8	10
10	6	3	6	4	9	10
24	8	5	4	5	13	9
4+6 *	4	4	4	5	8	9

All time-points are in weeks.

*Animals exposed during 4 weeks and then left to recover for 6 weeks.

For the follow-up experiments, animals were acutely exposed only to (R)-carvone, to heptanal alone, or to the combination of both. The final concentration of each odorant was 1mM. The odorants were directly added to the water bottles, without dilution in mineral oil. Therefore, the controls were kept with pure water. The water bottles were changed twice a week. The exposure started from at least E16.5 and the WOM was collected at 10 weeks of age. For each group, 3 males and 3 females were used.

qRT-PCR expression estimation.

For qRT-PCR experiments, RNA from WOM was extracted as previously described. 1 μg of 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 Technologies) following the manufacturer's instructions. Mean cycle threshold (Ct) values were obtained from two technical replicates, each normalised 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 t-test, with multiple-testing correction by the Benjamini & Hochberg (FDR) method.

Appendix B

Supplementary tables

sample	strain	tissue	age	sex	Illumina platform	read length	stranded	ENA ID
Transcriptome analysis of the WOM and VNO of male and female mice – Chapter 2.								
male1	C57BL/6J	VNO	8-10 weeks	male	Genome Analyzer II	76	no	ERS092040
male2	C57BL/6J	VNO	8-10 weeks	male	Genome Analyzer II	76	no	ERS092041
male3	C57BL/6J	VNO	8-10 weeks	male	Genome Analyzer II	76	no	ERS092042
female1	C57BL/6J	VNO	8-10 weeks	female	Genome Analyzer II	76	no	ERS092043
female2	C57BL/6J	VNO	8-10 weeks	female	Genome Analyzer II	76	no	ERS092044
female3	C57BL/6J	VNO	8-10 weeks	female	Genome Analyzer II	76	no	ERS092045
male1	C57BL/6J	WOM	8-10 weeks	male	HiSeq 2000	76	no	ERS092545
male2	C57BL/6J	WOM	8-10 weeks	male	HiSeq 2000	76	no	ERS092547
male3	C57BL/6J	WOM	8-10 weeks	male	HiSeq 2000	76	no	ERS092549
female1	C57BL/6J	WOM	8-10 weeks	female	HiSeq 2000	76	no	ERS092546
female2	C57BL/6J	WOM	8-10 weeks	female	HiSeq 2000	76	no	ERS092548
female3	C57BL/6J	WOM	8-10 weeks	female	HiSeq 2000	76	no	ERS092550

sample	strain	tissue	age	sex	Illumina platform	read length	stranded	ENA ID
RNAseq of mice lacking a cluster of OR genes in chromosome 9 – Chapter 2.								
delta1	129/SvEv-ΔOlfr7Δ	WOM	9 weeks	male	HiSeq 2500	100	yes	ERS473426
delta2	129/SvEv-ΔOlfr7Δ	WOM	9 weeks	male	HiSeq 2500	100	yes	ERS473427
delta3	129/SvEv-ΔOlfr7Δ	WOM	9 weeks	male	HiSeq 2500	100	yes	ERS473428
Comparison of the transcriptome of the OSNs versus the WOM – Chapter 3.								
WOM1	OMP-GFP	WOM	21 days	male	HiSeq 2000	100	no	ERS252155
WOM2	OMP-GFP	WOM	21 days	male	HiSeq 2000	100	no	ERS252156
WOM3	OMP-GFP	WOM	21 days	female	HiSeq 2000	100	no	ERS252157
OSN1	OMP-GFP	FACS OSNs	25 days	mixed	HiSeq 2000	100	no	ERS252158
OSN2	OMP-GFP	FACS OSNs	25 days	mixed	HiSeq 2000	100	no	ERS252159
OSN3	OMP-GFP	FACS OSNs	25 days	mixed	HiSeq 2000	100	no	ERS252160
Characterisation of two subpopulations of OMP$^+$ OSNs – Chapter 3.								
GFP^{low}1	OMP-GFP	FACS OSNs	25 weeks	male	HiSeq 2500	100	no	ERS715983
GFP^{low}2	OMP-GFP	FACS OSNs	25 weeks	male	HiSeq 2500	100	no	ERS715985
GFP^{low}3	OMP-GFP	FACS OSNs	25 weeks	male	HiSeq 2500	100	no	ERS715987
GFP^{high}1	OMP-GFP	FACS OSNs	25 weeks	male	HiSeq 2500	100	no	ERS715984
GFP^{high}2	OMP-GFP	FACS OSNs	25 weeks	male	HiSeq 2500	100	no	ERS715986
GFP^{high}3	OMP-GFP	FACS OSNs	25 weeks	male	HiSeq 2500	100	no	ERS715988

sample	strain	tissue	age	sex	Illumina platform	read length	stranded	ENA ID
The transcriptome of single OSNs – Chapter 3.								
single OSNs	OMP-GFP	single OSNs	23 weeks	male	HiSeq 2500	100	no	See Table B.4 for details
Comparison of the OR expression profile in different strains of mice – Chapter 4.								
B6_1	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658588
B6_2	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658589
B6_3	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658590
B6_4	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658591
B6_5	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658592
B6_6	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658593
129_1	129S5/SvEv	WOM	11 weeks	male	HiSeq 2000	100	no	ERS215497
129_2	129S5/SvEv	WOM	11 weeks	male	HiSeq 2000	100	no	ERS215498
129_3	129S5/SvEv	WOM	11 weeks	male	HiSeq 2000	100	no	ERS215499
cast1	CAST/Ei	WOM	12 weeks	female	HiSeq 2500	100	yes	ERS473423
cast2	CAST/Ei	WOM	12 weeks	female	HiSeq 2500	100	yes	ERS473424
cast3	CAST/Ei	WOM	12 weeks	female	HiSeq 2500	100	yes	ERS473425
Dissecting the genetic from the environmental effects on OR gene expression – Chapter 4.								
black1	C57BL/6NTac	WOM	10 weeks	female	HiSeq 2000	100	no	ERS373470
black2	C57BL/6NTac	WOM	10 weeks	female	HiSeq 2000	100	no	ERS373471
black3	C57BL/6NTac	WOM	10 weeks	female	HiSeq 2000	100	no	ERS373472
black4	C57BL/6NTac	WOM	10 weeks	female	HiSeq 2000	100	no	ERS373473
black5	C57BL/6NTac	WOM	10 weeks	male	HiSeq 2000	100	no	ERS373474
black6	C57BL/6NTac	WOM	10 weeks	male	HiSeq 2000	100	no	ERS373475
agouti1	129S5/SvEv	WOM	10 weeks	female	HiSeq 2000	100	no	ERS373476
agouti2	129S5/SvEv	WOM	10 weeks	female	HiSeq 2000	100	no	ERS373477
agouti3	129S5/SvEv	WOM	10 weeks	female	HiSeq 2000	100	no	ERS373478
agouti4	129S5/SvEv	WOM	10 weeks	female	HiSeq 2000	100	no	ERS373479
agouti5	129S5/SvEv	WOM	10 weeks	male	HiSeq 2000	100	no	ERS373480
agouti6	129S5/SvEv	WOM	10 weeks	male	HiSeq 2000	100	no	ERS373481

sample	strain	tissue	age	sex	Illumina platform	read length	stranded	ENA ID
Transcriptome of the WOM of newborn mice – Chapter 4.								
pups1	C57BL/6J	WOM	E19.5	mixed	HiSeq 2000	100	no	ERS223116
pups2	C57BL/6J	WOM	E19.5	mixed	HiSeq 2000	100	no	ERS223117
pups3	C57BL/6J	WOM	E19.5	mixed	HiSeq 2000	100	no	ERS223118
OR expression after exposure to a mix of odorants – Chapter 5.								
control1	C57BL/6J	WOM	24 weeks	male	HiSeq 2500	100	yes	ERS427453
control2	C57BL/6J	WOM	24 weeks	male	HiSeq 2500	100	yes	ERS427454
control3	C57BL/6J	WOM	24 weeks	male	HiSeq 2500	100	yes	ERS427455
control4	C57BL/6J	WOM	24 weeks	female	HiSeq 2500	100	yes	ERS427456
control5	C57BL/6J	WOM	24 weeks	female	HiSeq 2500	100	yes	ERS427457
control6	C57BL/6J	WOM	24 weeks	female	HiSeq 2500	100	yes	ERS427458
odour1	C57BL/6J	WOM	24 weeks	male	HiSeq 2500	100	yes	ERS427447
odour2	C57BL/6J	WOM	24 weeks	male	HiSeq 2500	100	yes	ERS427448
odour3	C57BL/6J	WOM	24 weeks	male	HiSeq 2500	100	yes	ERS427449
odour4	C57BL/6J	WOM	24 weeks	female	HiSeq 2500	100	yes	ERS427450
odour5	C57BL/6J	WOM	24 weeks	female	HiSeq 2500	100	yes	ERS427451
odour6	C57BL/6J	WOM	24 weeks	female	HiSeq 2500	100	yes	ERS427452
OR expression after exposure to particular odorants – Chapter 5.								
carvone1	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658594
carvone2	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658595
carvone3	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658596
carvone4	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658597
carvone5	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658598
carvone6	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658599
heptanal1	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658600
heptanal2	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658601
heptanal3	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658602
heptanal4	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658603
heptanal5	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658604
heptanal6	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658605

sample	strain	tissue	age	sex	Illumina platform	read length	stranded	ENA ID
both1	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658606
both2	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658607
both3	C57BL/6J	WOM	10 weeks	male	HiSeq 2500	100	yes	ERS658608
both4	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658609
both5	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658610
both6	C57BL/6J	WOM	10 weeks	female	HiSeq 2500	100	yes	ERS658611

Table B.1 – Sequenced samples presented in this dissertation. Details about each of the samples used for RNAseq. All sequencing was paired-end; the read length is indicated, in basepairs. *Stranded* indicates whether the library preparation method was strand-specific or not. All raw data are available through the European Nucleotide Archive (ENA).

sample	total fragments	uniquely mapped %		multimapped %		unmapped %	
Transcriptome analysis of the VNO of male and female mice – Chapter 2.							
male1	33,829,828	27,791,186	82.15	2,356,980	6.97	3,681,662	10.88
male2	34,334,069	27,943,814	81.39	2,300,990	6.70	4,089,265	11.91
male3	33,452,308	26,979,727	80.65	2,259,065	6.75	4,213,516	12.60
female1	38,989,649	30,690,761	78.72	2,517,113	6.46	5,781,775	14.83
female2	41,267,287	33,471,377	81.11	1,828,650	4.43	5,967,260	14.46
female3	40,783,743	33,330,682	81.73	2,907,635	7.13	4,545,426	11.15
Transcriptome analysis of the WOM of male and female mice – Chapter 2.							
male1	47,449,378	43,428,430	91.53	2,422,702	5.11	1,598,246	3.37
male2	45,919,675	41,968,773	91.40	2,815,735	6.13	1,135,167	2.47
male3	45,436,958	38,304,453	84.30	5,906,163	13.00	1,226,342	2.70
female1	41,096,169	35,868,924	87.28	4,003,075	9.74	1,224,170	2.98
female2	53,985,044	46,021,315	85.25	6,361,505	11.78	1,602,224	2.97
female3	44,548,659	38,716,506	86.91	4,838,407	10.86	993,746	2.23
RNAseq of mice lacking a cluster of OR genes in chromosome 9 – Chapter 2.							
delta1	40,815,069	37,230,201	91.22	2,428,922	5.95	1,155,946	2.83
delta2	41,774,414	38,165,713	91.36	2,444,901	5.85	1,163,800	2.79
delta3	48,779,436	44,509,100	91.25	2,975,024	6.10	1,295,312	2.66
Comparison of the transcriptome of the OSNs versus the WOM – Chapter 3.							
WOM1	43,534,928	38,820,863	89.17	1,723,998	3.96	2,990,067	6.87
WOM2	75,289,455	67,690,537	89.91	3,018,346	4.01	4,580,572	6.08
WOM3	54,231,767	49,952,440	92.11	2,316,878	4.27	1,962,449	3.62
OSN1	48,523,309	45,373,409	93.51	1,764,633	3.64	1,385,267	2.85
OSN2	57,565,818	46,820,001	81.33	2,142,656	3.72	8,603,161	14.94
OSN3	75,288,647	69,454,921	92.25	2,506,461	3.33	3,327,265	4.42
Characterisation of two subpopulations of OMP$^+$ OSNs – Chapter 3.							
GFP^{low}1	66,274,523	58,952,254	88.95	3,619,369	5.46	3,702,900	5.59
GFP^{low}2	66,293,232	59,198,484	89.30	3,343,118	5.04	3,751,630	5.66
GFP^{low}3	80,748,448	72,241,026	89.46	4,113,321	5.09	4,394,101	5.44
GFP^{high}1	17,734,782	15,059,594	84.92	908,067	5.12	1,767,121	9.96
GFP^{high}2	72,410,349	64,410,818	88.95	3,528,935	4.87	4,470,596	6.17
GFP^{high}3	73,843,358	65,957,173	89.32	3,764,830	5.10	4,121,355	5.58

sample	total fragments	uniquely mapped %		multimapped %		unmapped %	
Comparison of the OR expression profile in different strains of mice – Chapter 4.							
B6_1	37,332,765	32,250,593	86.39	2,121,241	5.68	5,082,164	7.93
B6_2	58,184,940	51,710,263	88.87	3,191,347	5.48	6,474,671	5.65
B6_3	46,459,677	40,920,392	88.08	3,338,396	7.19	5,539,280	4.74
B6_4	39,423,479	34,897,650	88.52	2,466,335	6.26	4,525,824	5.22
B6_5	24,228,124	20,190,426	83.33	1,468,759	6.06	4,037,687	10.61
B6_6	36,811,932	32,155,493	87.35	2„712,152	7.37	4,656,434	5.28
129_1	51,360,567	41,641,267	81.08	6,228,439	12.13	3,490,861	6.80
129_2	56,018,117	49,911,723	89.10	2,897,953	5.17	3,208,441	5.73
129_3	75,872,597	68,102,703	89.76	4,015,197	5.29	3,754,697	4.95
cast1	42,193,697	38,185,627	90.50	2,144,719	5.08	1,863,351	4.42
cast2	35,534,499	31,307,518	88.10	2,553,077	7.18	1,673,904	4.71
cast3	46,273,696	41,504,133	89.69	2,618,676	5.66	21,50,887	4.65
Dissecting the genetic from the environmental effects on OR gene expression – Chapter 4.							
black1	53,532,994	45,079,303	84.21	5,431,071	10.15	3,022,620	5.65
black2	74,253,096	64,338,991	86.65	5,185,893	6.98	4,728,212	6.37
black3	41,608,225	37,543,713	90.23	1,937,432	4.66	2,127,080	5.11
black4	71,212,832	63,952,008	89.80	3,387,536	4.76	3,873,288	5.44
black5	51,920,894	45,650,480	87.92	3,106,416	5.98	3,163,998	6.09
black6	90,279,406	77,481,328	85.82	7,415,649	8.21	5,382,429	5.96
agouti1	60,959,853	53,430,862	87.65	3,088,901	5.07	4,440,090	7.28
agouti2	26,709,804	23,547,335	88.16	1,286,307	4.82	1,876,162	7.02
agouti3	30,791,098	26,913,940	87.41	1,483,789	4.82	2,393,369	7.77
agouti4	49,844,784	43,363,086	87.00	3,638,676	7.30	2,843,022	5.70
agouti5	34,387,223	29,030,014	84.42	3,412,419	9.92	1,944,790	5.66
agouti6	41,895,931	37,528,206	89.57	1,927,666	4.60	2,440,059	5.82
Transcriptome of the WOM of newborn mice – Chapter 4.							
pups1	49,335,880	41,680,801	84.48	3,860,509	7.82	3,794,570	7.69
pups2	40,532,710	35,680,710	88.03	2,830,765	6.98	2,021,235	4.99
pups3	64,224,553	57,477,476	89.49	4,100,707	6.38	2,646,370	4.12
OR expression after exposure to a mix of odorants – Chapter 5.							
control1	52,160,507	46,117,696	88.41	4,374,645	8.39	1,668,166	3.20

sample	total fragments	uniquely mapped %		multimapped %		unmapped %	
control2	45,667,031	41,594,467	91.08	2,571,602	5.63	1,500,962	3.29
control3	45,665,776	41,565,777	91.02	2,501,129	5.48	1,598,870	3.50
control4	54,725,715	49,969,350	91.31	3,109,856	5.68	1,646,509	3.01
control5	46,906,572	42,584,649	90.79	2,784,963	5.94	1,536,960	3.28
control6	51,235,209	45,397,000	88.61	4,063,968	7.93	1,774,241	3.46
odour1	53,005,866	48,477,141	91.46	2,796,147	5.28	1,732,578	3.27
odour2	44,239,992	39,344,128	88.93	3,379,210	7.64	1,516,654	3.43
odour3	50,470,024	45,624,241	90.40	3,155,133	6.25	1,690,650	3.35
odour4	48,495,642	43,996,672	90.72	2,870,256	5.92	1,628,714	3.36
odour5	50,189,225	43,976,059	87.62	4,345,171	8.66	1,867,995	3.72
odour6	50,338,265	45,770,319	90.93	2,824,275	5.61	1,743,671	3.46
OR expression after exposure to particular odorants – Chapter 5.							
carvone1	34,538,910	30,563,297	88.49	1,892,137	5.48	2,083,476	6.03
carvone2	35,531,736	31,117,642	87.58	1,800,586	5.07	2,613,508	7.36
carvone3	32,064,615	27,022,540	84.28	3,134,406	9.78	1,907,669	5.95
carvone4	33,834,834	29,883,589	88.32	1,984,988	5.87	1,966,257	5.81
carvone5	41,188,840	36,500,265	88.62	2,551,702	6.20	2,136,873	5.19
carvone6	33,179,966	28,071,868	84.60	3,700,369	11.15	1,407,729	4.24
heptanal1	36,368,912	30,577,330	84.08	3,670,436	10.09	2,121,146	5.83
heptanal2	47,799,810	42,578,959	89.08	2,691,320	5.63	2,529,531	5.29
heptanal3	71,528,814	63,800,289	89.20	3,801,877	5.32	3,926,648	5.49
heptanal4	34,423,907	31,050,666	90.20	1,816,408	5.28	1,556,833	4.52
heptanal5	39,095,683	34,761,037	88.91	2,190,893	5.60	2,143,753	5.48
heptanal6	14,901,039	12,886,192	86.48	1,105,993	7.42	908,854	6.10
both1	44,818,699	39,855,390	88.93	2,429,371	5.42	2,533,938	5.65
both2	37,287,128	33,330,936	89.39	2,084,171	5.59	1,872,021	5.02
both3	43,729,818	31,058,026	71.02	3,500,396	8.00	9,171,396	20.97
both4	50,876,569	45,689,933	89.81	2,696,457	5.30	2,490,179	4.89
both5	35,849,711	32,049,662	89.40	1,908,123	5.32	1,891,926	5.28
both6	41,060,091	35,784,547	87.15	3,004,507	7.32	2,271,037	5.53

Table B.2 – Mapping statistics of RNAseq samples. Mapping statistics of the samples sequenced (see also Table B.1).

Ensembl ID	Ensembl gene name	Matched cDNA from [65,136]
ENSMUSG00000096294	*Gm10302*	*Vmn2r47*
ENSMUSG00000096871	*Gm10665*	*Vmn1r102*
ENSMUSG00000096348	*Gm10666*	*Vmn1r141.Vmn1r93*
ENSMUSG00000094762	*Gm10670*	*Vmn1r150*
ENSMUSG00000087688	*Gm11300*	*Vmn1r203*
ENSMUSG00000087643	*Gm11314*	*Vmn1r208*
ENSMUSG00000096152	*Gm16442*	*Vmn1r140*
ENSMUSG00000095745	*Gm4133*	*Vmn1r146*
ENSMUSG00000095837	*Gm4141*	*Vmn1r106*
ENSMUSG00000093941	*Gm4172*	*Vmn1r131*
ENSMUSG00000096513	*Gm4175*	*Vmn1r133*
ENSMUSG00000096760	*Gm4177*	*Vmn1r134*
ENSMUSG00000095163	*Gm4179*	*Vmn1r138*
ENSMUSG00000093871	*Gm4187*	*Vmn1r98*
ENSMUSG00000095984	*Gm4201*	*Vmn1r154*
ENSMUSG00000092297	*Gm4214*	*Vmn1r161*
ENSMUSG00000094532	*Gm4216*	*Vmn1r162*
ENSMUSG00000096073	*Gm4220*	*Vmn1r166*
ENSMUSG00000094757	*Gm4498*	*Vmn1r145*
ENSMUSG00000095191	*Gm5725*	*Vmn1r136*
ENSMUSG00000096761	*Gm5726*	*Vmn1r105*
ENSMUSG00000095806	*Gm5728*	*Vmn1r147*
ENSMUSG00000094298	*Gm6164*	*Vmn1r144*
ENSMUSG00000094149	*Gm8453*	*Vmn1r97*
ENSMUSG00000094981	*Gm8653*	*Vmn1r96*
ENSMUSG00000093917	*Gm8660*	*Vmn1r99*
ENSMUSG00000094748	*Gm8677*	*Vmn1r153*
ENSMUSG00000095081	*Gm8693*	*Vmn1r108.Vmn1r156*
ENSMUSG00000096601	*Gm8720*	*Vmn1r164*
ENSMUSG00000091528	*Gm9268*	*Vmn2r64*
ENSMUSG00000096304	*RP23-331M13.5*	*Vmn1r92*
ENSMUSG00000092456	*V1rd19*	*Vmn1r182*

Table B.3 – VR genes not properly annotated in Ensembl. The matched cDNA sequences are those that aligned with 100% coverage and 100% identity, indicating that they represent the same gene but haven't been properly annotated in Ensembl. Other genes matched VR sequences with lower identity and most likely represent unannotated paralogs, but were not included in the analyses since there is a lack of annotation for them.

sample	total fragments	unique	%	multimapped	%	unmapped	%	included	ENA ID
OSN_171	5,403,186	4,587,828	84.91	296,854	5.49	518,504	9.59	yes	ERS361292
OSN_177	3,416,492	2,962,593	86.71	171,355	5.02	282,544	8.27	yes	ERS361298
OSN_183	4,884,518	4,139,987	84.76	259,317	5.31	485,214	9.93	yes	ERS361304
OSN_188	4,087,523	3,468,899	84.87	235,054	5.75	383,570	9.38	yes	ERS361309
OSN_193	3,490,923	2,908,510	83.32	164,005	4.70	418,408	11.98	yes	ERS361314
OSN_195	3,376,521	2,496,006	73.92	390,214	11.56	490,301	14.52	yes	ERS361316
OSN_201	4,604,541	4,055,065	88.07	185,030	4.02	364,446	7.91	yes	ERS361322
OSN_204	4,187,094	3,587,283	85.67	204,236	4.88	395,575	9.45	yes	ERS361325
OSN_205	5,487,975	4,787,155	87.23	260,788	4.75	440,032	8.02	yes	ERS361326
OSN_216	4,805,114	4,155,706	86.49	257,473	5.36	391,935	8.16	yes	ERS361337
OSN_222	4,080,624	3,481,018	85.31	241,642	5.92	357,964	8.77	yes	ERS361343
OSN_224	3,370,232	2,723,513	80.81	149,548	4.44	497,171	14.75	yes	ERS361345
OSN_230	4,138,735	3,379,738	81.66	247,198	5.97	511,799	12.37	yes	ERS361351
OSN_236	2,962,912	2,467,855	83.29	104,321	3.52	390,736	13.19	yes	ERS361357
OSN_238	3,633,203	3,058,970	84.19	214,892	5.91	359,341	9.89	yes	ERS361359
OSN_243	5,146,808	4,440,478	86.28	269,240	5.23	437,090	8.5	yes	ERS361364
OSN_251	5,069,051	4,216,080	83.17	217,145	4.28	635,826	12.54	yes	ERS361372
OSN_259	4,997,202	4,331,122	86.67	197,519	3.95	468,561	9.38	yes	ERS361380
OSN_261	6,936,460	6,092,355	87.83	297,816	4.29	546,289	7.87	yes	ERS361382
OSN_262	4,420,237	3,770,751	85.31	277,302	6.27	372,184	8.42	yes	ERS361383
OSN_263	5,688,875	4,560,119	80.16	399,782	7.03	728,974	12.81	yes	ERS361384
OSN_178	3,078,169	2,644,556	85.91	165,023	5.36	268,590	8.73	no	ERS361299
OSN_185	4,404,713	3,789,209	86.03	240,960	5.47	374,544	8.51	no	ERS361306
OSN_191	4,136,140	3,487,140	84.31	285,589	6.90	363,411	8.79	no	ERS361312
OSN_207	4,378,400	3,830,209	87.48	175,419	4.01	372,772	8.52	no	ERS361328
OSN_214	4,120,952	3,398,111	82.46	270,927	6.57	451,914	10.96	no	ERS361335
OSN_218	4,693,293	3,890,545	82.90	242,345	5.16	560,403	11.94	no	ERS361339
OSN_223	3,897,470	3,319,616	85.17	185,317	4.75	392,537	10.07	no	ERS361344
OSN_255	4,790,253	3,705,418	77.35	449,852	9.39	634,983	13.25	no	ERS361376
OSN_257	4,911,385	4,252,816	86.59	172,097	3.50	486,472	9.91	no	ERS361378

Table B.4 – Mapping statistics of RNAseq single-OSN samples. Mapping statistics of the single-OSN samples sequenced. Column *included* indicates whether the sample was included in downstream analyses after the QC stage. Excluded samples showed expression of more than a single abundant OR gene and represent carry-over from adjacent wells or could contain two cells. All raw data is available through the European Nucleotide Archive (ENA).

Appendix C

Papers produced during my PhD.

C.1 Papers associated with this dissertation.

- Ibarra-Soria, X., Levitin, M. O. & Logan, D. W. The genomic basis of vomeronasal-mediated behaviour. *Mamm Genome* **25**, 75–86 (2014). DOI: 10.1007/s00335-013-9463-1.

- Ibarra-Soria, X., Levitin, M. O., Saraiva, L. R. & Logan, D. W. The olfactory transcriptomes of mice. *PLoS Genet* **10**, e1004593 (2014). DOI: 10.1371/journal.pgen.1004593.

- Saraiva, L. R.*, Ibarra-Soria, X.*, Khan, M., Omura, M., Scialdone, A., Mombaerts, P., Marioni, J. C. & Logan, D. W. Hierarchical deconstruction of mouse olfactory sensory neurons: from whole mucosa to single-cell RNA-seq. *Sci Rep* **5**, 18178 (2015). DOI: 10.1038/srep18178.

  * Contributed equally.

C.2 Other papers.

- Dey, S., Chamero, P., Pru, J. K., Chien, M.-S., Ibarra-Soria, X., Spencer, K. R., Logan, D. W., Matsunami, H., Peluso, J. J., & Stowers, L. Cyclic regulation of sensory perception by a female hormone alters behavior. *Cell* **161**, 1334–1344 (2015). DOI: 10.1016/j.cell.2015.04.052.

- Oboti, L., Ibarra-Soria, X., Pérez-Gómez, A., Schmid, A., Pyrski, M., Paschek, N., Kircher, S., Logan, D. W., Leinders-Zufall, T., Zufall, F. & Chamero, P. Pregnancy

and estrogen enhance neural progenitor-cell proliferation in the vomeronasal sensory epithelium. *BMC Biology* **13**, 104 (2015). DOI: 10.1186/s12915-015-0211-8.

- Nakahara, T. S., Cardozo, L. M., Ibarra-Soria, X., Bard, A., Carvalho, V. M. A., Trintinalia, G. Z. , Logan, D. W., & Papes, F., Detection of pup odors by adult vomeronasal neurons non-canonically expressing an odorant receptor gene is influenced by sex and parenting status, *BMC Biology*, **In press**.