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

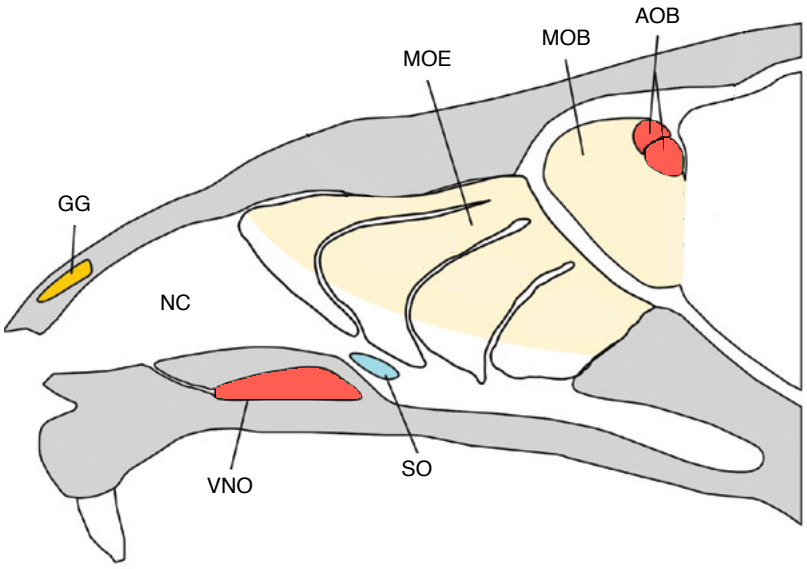

Figure 1.1 – The mammalian olfactory system. Schematic of the mouse nose (sagittal view) and the different components of the olfactory system. At the back of the nasal cavity (NC) is the main olfactory epithelium (MOE). Directly above the roof of the mouth, at the base of the nasal cavity, is the vomeronasal 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).

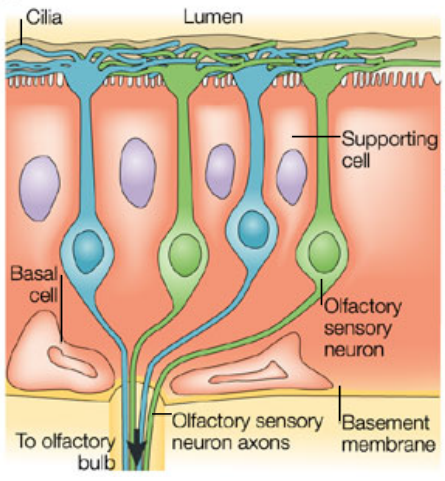

Figure 1.2 – Composition of the MOE. Schematic representing the different cell types present in the main olfactory epithelium and their organisation. In blue and green are the bipolar olfactory sensory neurones (OSNs), that extend dendrites towards the lumen of the MOE; dendrites terminate in long cilia 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).

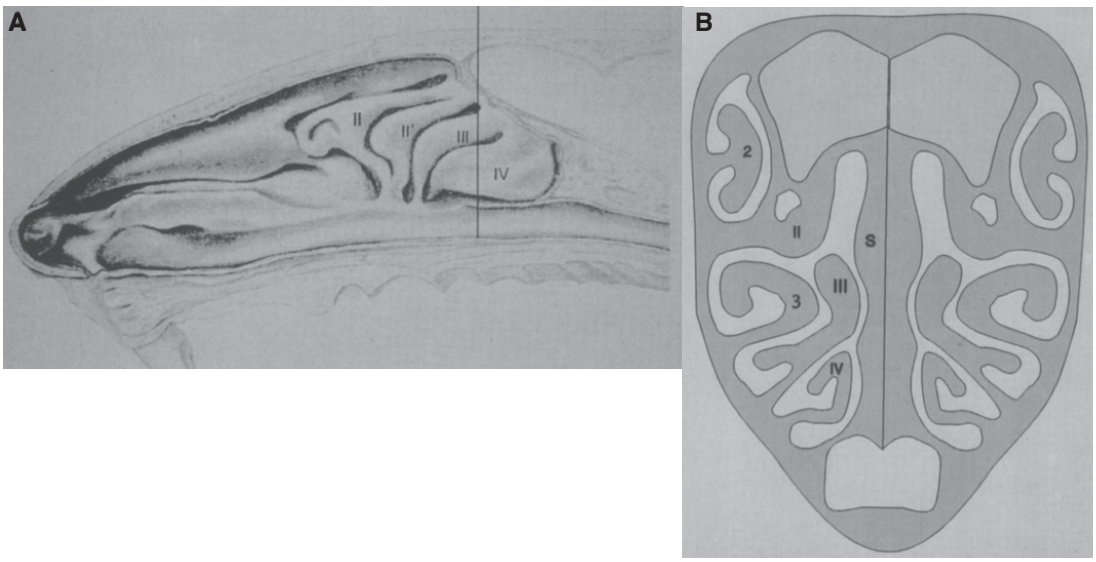

Figure 1.3 – **Turbinate structure of the MOE. A)** Sagittal view of the mouse nose. The vertical line indicates the plane of section represented in **B)** in a coronal view. The numerals indicate the individual turbinates. II, II', III and IV endoturbinates; 2, 3 ectoturbinates. s, nasal septum. Reproduced 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

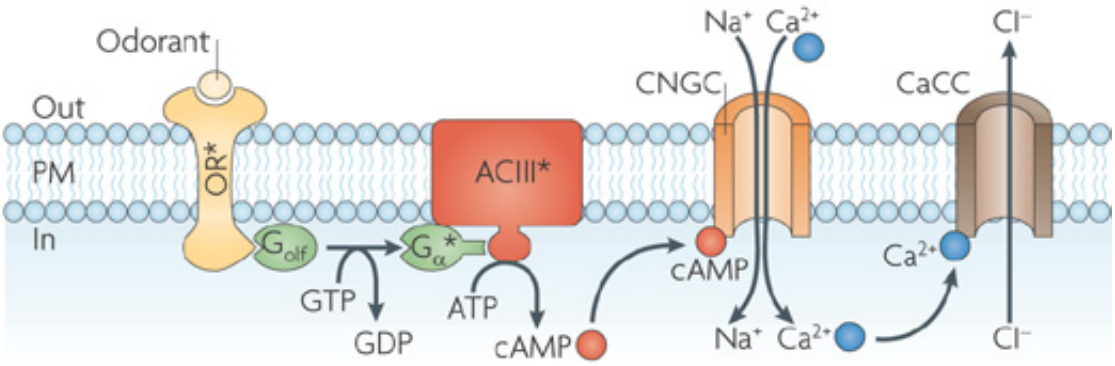

Figure 1.4 – Olfactory signal transduction cascade. When an odorant binds to an olfactory receptor (OR) it activates the trimeric G protein $G\alpha_{olf}$ which in turns activates the adenylate cyclase ACIII. ACIII catalyses the conversion of ATP into cAMP, which serves as a second messenger to open 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;

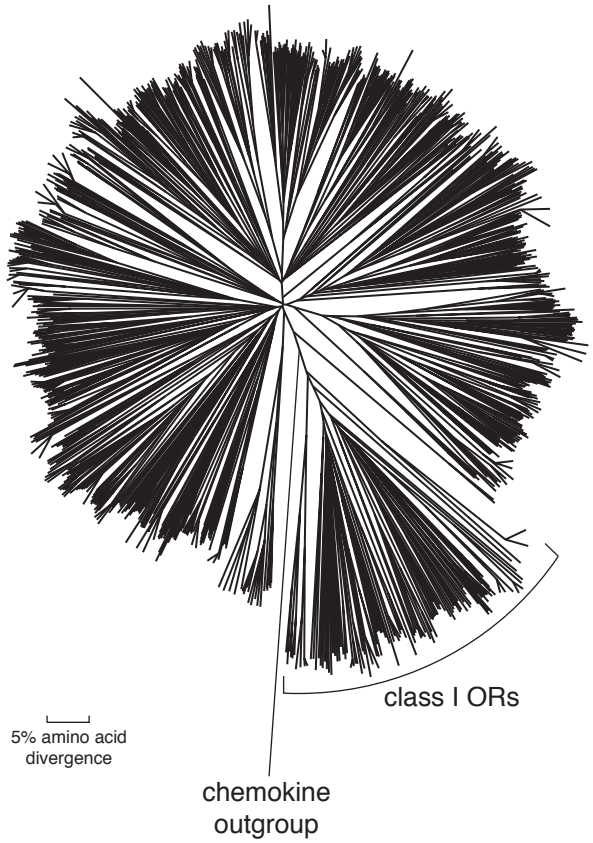

Figure 1.5 – The mouse olfactory receptor gene family. Phylogenetic tree of the mouse olfactory receptor genes. The class I genes are indicated and the remaining are class II. Reproduced from [62].

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].

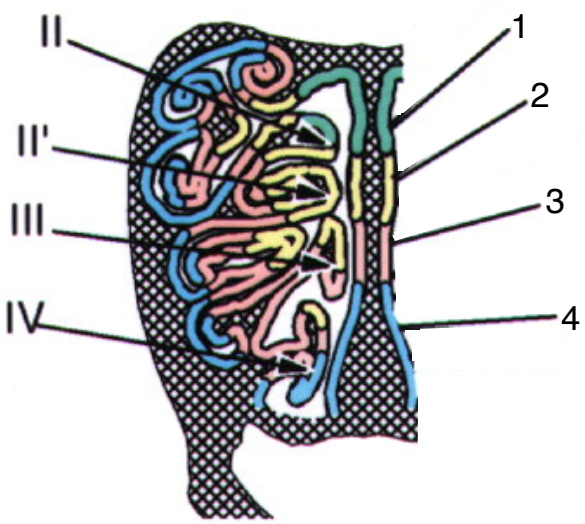

Figure 1.6 – Olfactory receptors are expressed in zones. Schematic of a coronal section of the MOE. The endoturbinates are indicated. The zones 1 to 4 are delineated by different colours. Figure reproduced from [77] by permission of Oxford University Press.

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,

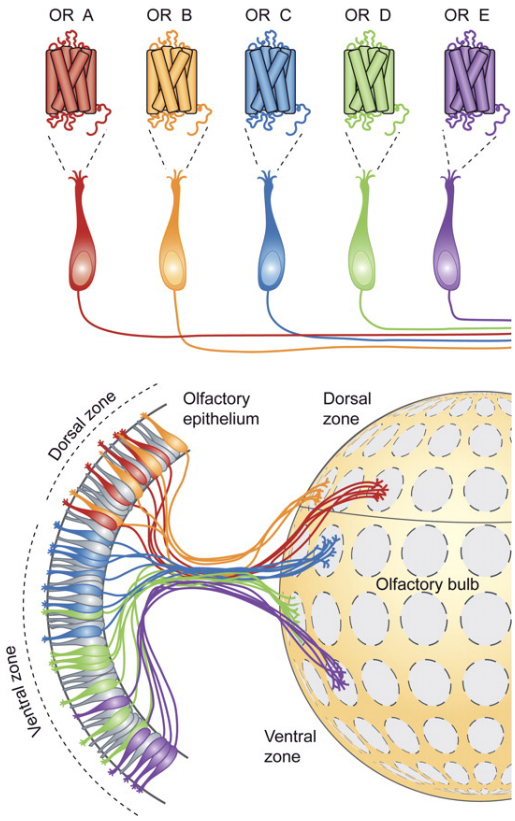

Figure 1.7 – OSN axons coalesce into glomeruli. Schematic representation of five different ORs; each patterns a specific subpopulation of OSNs, as indicated by the corresponding colour. The axons from all the OSNs expressing the same OR coalesce into a particular glomerulus in the main olfactory bulb.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

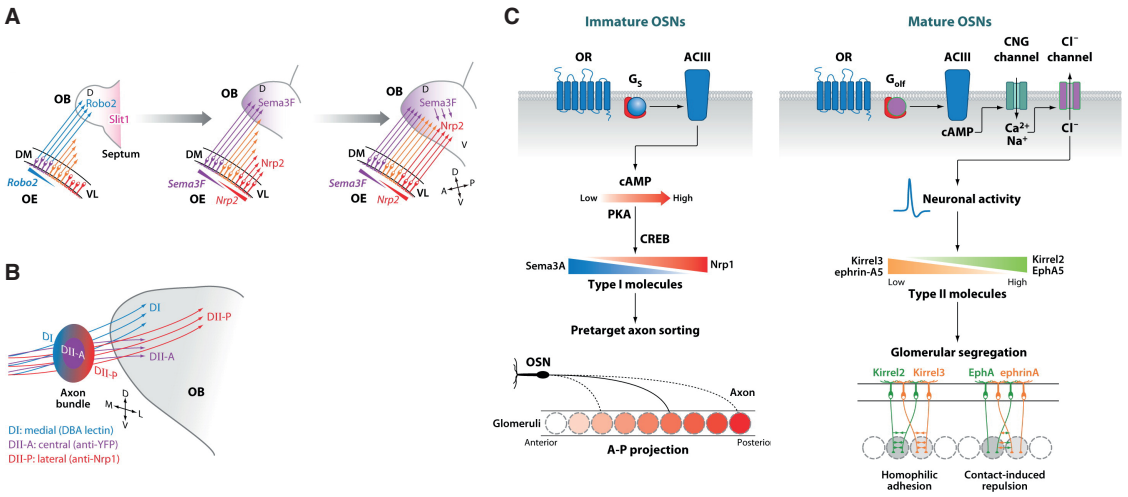

Figure 1.8 – Glomerular organisation in the olfactory bulb. A) A model for the DV projection of OSN axons. The development of the MOB starts from the dorsal domain and extends ventrally. Thus, axons from OSNs from the dorsal MOE reach the bulb first. *(Left)* These express Robo2 which is repelled by 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).

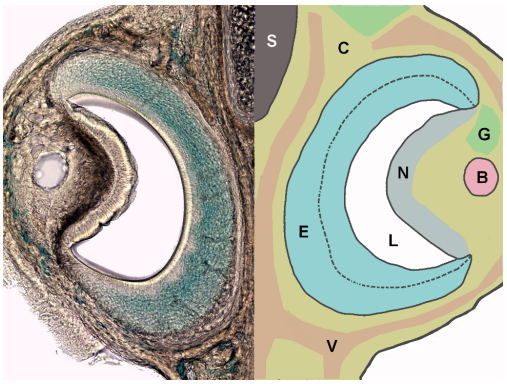

Figure 1.9 – The mouse vomeronasal organ. A coronal section through half of the VNO of an adult mouse *(left)* with a cartoon of the corresponding tissue morphology *(right)*. S, nasal septum; C, cavernous tissue; G, glandular tissue; B, blood vessel; V, vomer; N, nonsensory epithelium; L, lumen; E, 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,

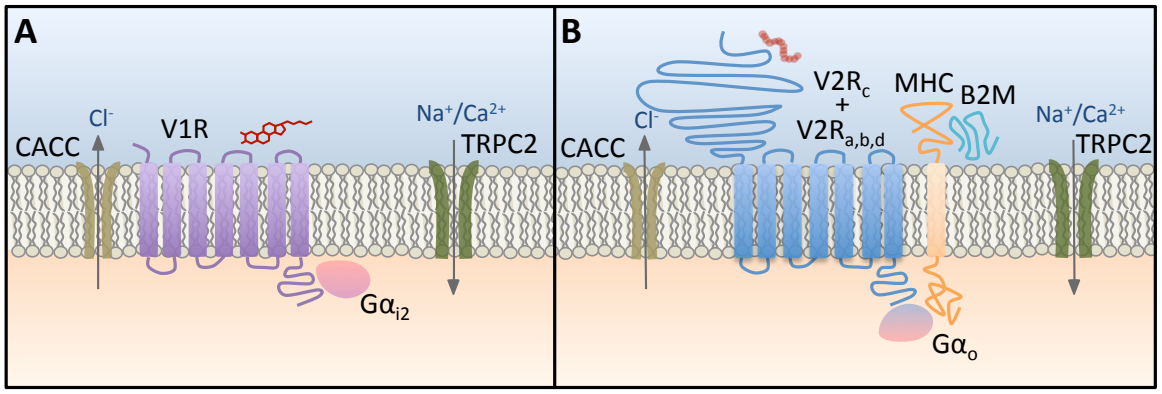

Figure 1.10 – Signal transduction proteins in vomeronasal sensory neurons. There are two subclasses of mammalian vomeronasal sensory neurones (VSNs). **A)** In apical VSNs, a V1R receptor associated with the $G\alpha_{i2}$ G-protein subunit is activated by a small, volatile chemical ligand. **B)** In 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*

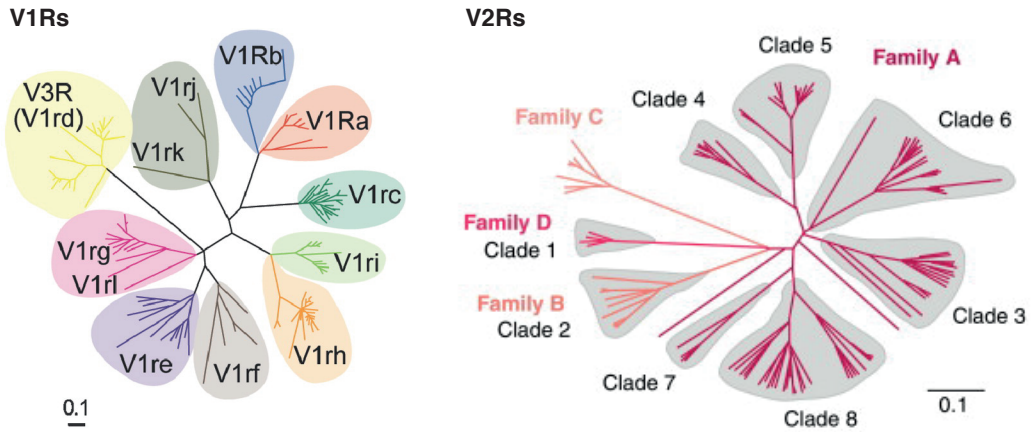

Figure 1.11 – The mouse vomeronasal receptor gene family. Phylogenetic trees of the V1R *(left)* and V2R *(right)* gene families. V1R tree reproduced from [144]. V2R tree reprinted from [145], copyright (2012) with permission from Elsevier.

[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

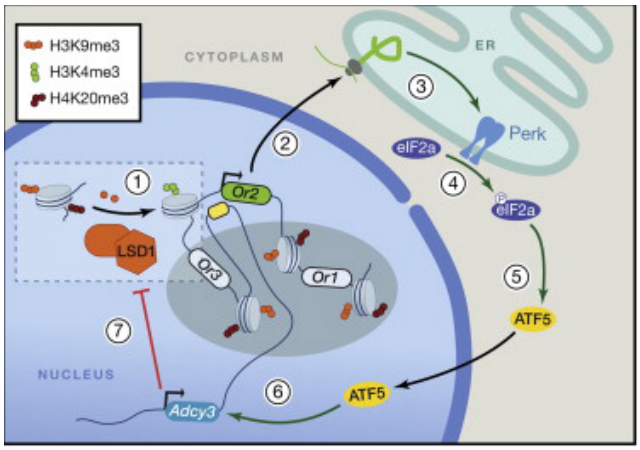

Figure 1.12 – A feedback mechanism ensures singular OR expression. Initially, the OR repertoire is silenced by condensation in foci (dark area in the nucleus) and marking with heterochromatin histone modifications (H3K9me3 and H4K20me3). (1) To activate OR expression, a single OR gene is desilenced by 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].

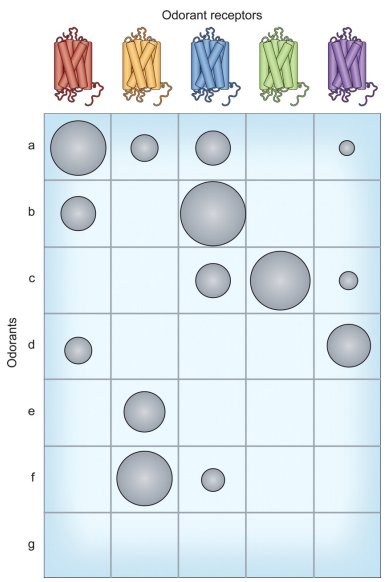

Figure 1.13 – **Combinatorial odour coding.** At the top are represented five different ORs. Below is a table of the responses of each OR to seven different odorants (a to g). The diameter of the circle indicates the magnitude of the response. Some odorants activate many receptors while others activate 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

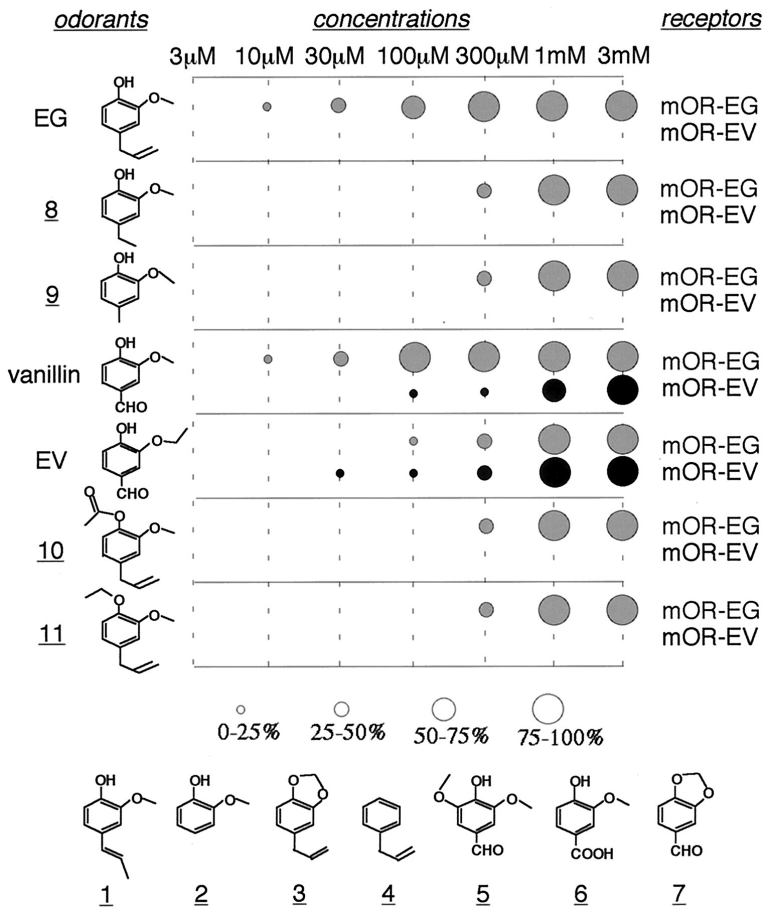

Figure 1.14 – Molecular range of ligands for the mOR-EG and mOR-EV receptors. Two mouse OR genes (mOR-EG and mOR-EV) were tested with a panel of odorants (at a range of concentrations) that are variations on the structure of eugenol and vanillin, the ligands of these receptors. Components 1 to 7 did 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-

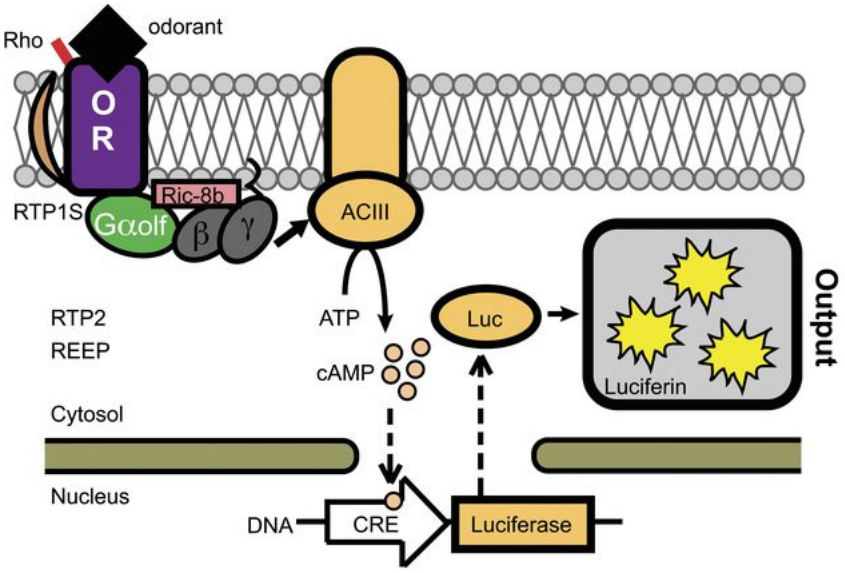

Figure 1.15 – *In vitro* expression of ORs in Hana3A cells. HEK293 cells stably expressing $G\alpha_{\text{olf}}$, RTP1S, RTP2 and REEP1, called Hana3A cells, are able to express a construct containing an OR gene and translocate it to the membrane with better efficiency than other systems. A luciferase 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-

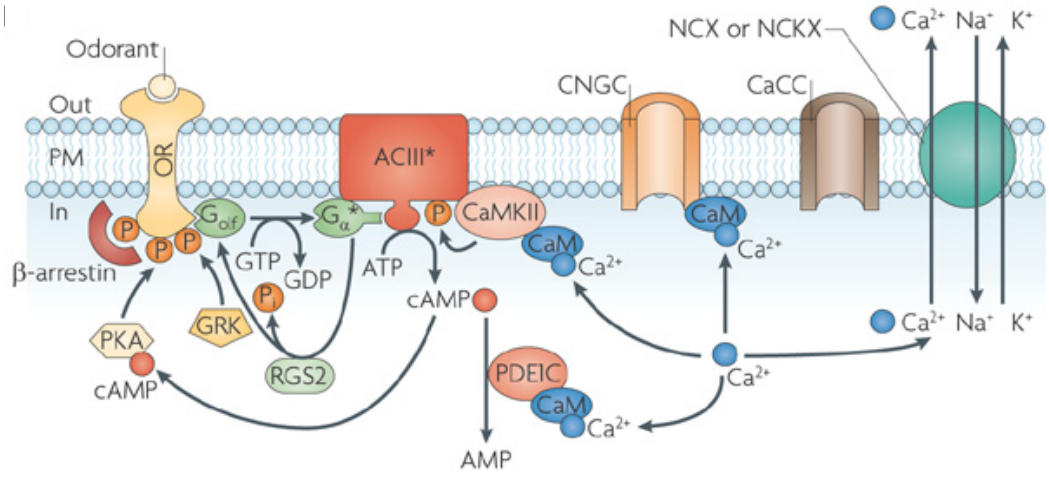

Figure 1.16 – Adaptation of OSNs to repeated stimulation. Upon stimulation of an OSN, Ca^{2+} influx induces adaptation by regulating the activity of key players in the signal transduction pathway. Ca^{2+} bound to calmodulin (CaM) interacts with the CNG channel and diminishes its affinity for cAMP to 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.