Human olfaction: from genomic variation to phenotypic diversity

Yehudit Hasin-Brunshtein, Doron Lancet and Tsviya Olender

Department of Molecular Genetics and the Crown Human Genome Center, Weizmann Institute of Science, Rehovot 76100, Israel

The sense of smell is a complex molecular device, encompassing several hundred olfactory receptor proteins (ORs). These receptors, encoded by the largest human gene superfamily, integrate odorant signals into an accurate ‘odor image’ in the brain. Widespread phenotypic diversity in human olfaction is, in part, attributable to prevalent genetic variation in OR genes, owing to copy number variation, deletion alleles and deleterious single nucleotide polymorphisms. The development of new genomic tools, including next generation sequencing and CNV assays, provides opportunities to characterize the genetic variations of this system. The advent of large-scale functional screens of expressed ORs, combined with genetic association studies, has the potential to link variations in ORs to human chemosensory phenotypes. This promises to provide a genome-wide view of human olfaction, resulting in a deeper understanding of personalized odor coding, with the potential to decipher flavor and fragrance preferences.

Understanding the olfactory molecular universe

Olfaction – the sense of smell – is a molecularly complex sensory processing system. It is capable of producing accurate odor perception, based on input from hundreds of sensory neuronal types equipped with diverse molecular sensors – olfactory receptor (OR) proteins. Olfaction is characterized by a remarkable ability to detect and discriminate thousands of low molecular mass compounds (odorants). Most organisms rely on olfactory cues for a wide range of activities, such as food acquisition, reproduction, migration and predator alarm. A comprehensive molecular understanding of this sensory pathway is now beginning to emerge through studies of inter-individual differences in smell sensitivity phenotypes and the elucidation of the relevant genomic diversity of OR genotypes.

The sense of smell involves a cascade of biochemical and electrophysiological processes, which help convert the molecular information of an odorant into odor sensation. The human olfactory epithelium accommodates ~10^7 olfactory sensory neurons, the ciliated dendritic ends of which receive and amplify chemosensory signals. Electrical impulses, which are subsequently generated in the axons of these neurons, are transmitted to synaptic complexes (glomeruli) of the olfactory bulb in the brain. The sensory neurons largely follow a ‘one neuron–one receptor’ rule, whereby each sensory cell randomly expresses only one out of all possible OR genes in a mono-allelic fashion. Each olfactory bulb glomerulus receives the input of a subgroup of neurons, all expressing the same OR protein. Consequently, integrating over all sensory neurons and glomeruli, an odorant is represented by a unique combination of activated units. A plethora of neurodevelopmental studies has resulted in a global view of how this intriguing wiring is made possible. Now, human genetics, together with cutting-edge genomics, are promising to shed further light on how odorants are deciphered.

Genome evolution of the OR repertoire: receptor birth and death

Key components of the molecular decoding device of the nose are OR proteins, belonging to the hyperfamily of seven-helix G-protein-coupled receptors (GPCRs), which are transducers of a wide array of extracellular molecular signals. ORs, like visual opsins, bitter taste receptors (T2Rs) and vomeronasal receptors (V1Rs), belong to the GPCR superfamily, and are characterized by a relatively compact helix–loop structure. A design principle common to ORs and other similar families of sensory receptors is amino acid variability in defined functional sequence positions, in conjunction with conserved sequence motifs that probably underlie structural integrity and interactions with downstream transduction partners. Variability among OR genes at the hypothetical odorant-binding site is thought to facilitate the recognition of diverse odorant ligands. In parallel, if two individuals possess distinct subsets of the OR gene repertoire of the human species, they might perceive odorants differently. This molecularly based sensory idiosyncrasy is the major focus of our review.

Like several other types of GPCRs, ORs have an intronless coding region, typically comprising ~1 kb of sequence. In contrast to fish ORs or mammalian vomeronasal receptors (chemosensory receptors of the vertebrate ‘alternative nose’), mammalian ORs show sufficient sequence motif conservation to enable facile member identification and cloning of the entire 800–2000-strong repertoire. This extremely large gene inventory has been classified by several computational approaches; for more details, see http://bioportal.weizmann.ac.il/HORDE/ and http://www.genenames.org/. Human ORs are organized in several dozen genomic clusters distributed on most chromosomes. This organization is highly conserved among mammals, including marsupials. A massive process of gene and cluster duplication resulted in a ~2-fold increase of the OR repertoire after the divergence from monotremes.
Box 1. Mammalian olfactory and vomeronasal receptor gene families

OR genes are present in practically all multicellular organisms. Fish and fly have only ~100 OR genes, whereas most mammals have >1000. Surprisingly, the simple nematode Caenorhabditis elegans has as many ORs as a mammal, indicating a universal underlying principle. Phylogenetic analyses have indicated that the huge tetrapod OR gene superfamily has originated from two genes inferred to be present in the recent common ancestor of fish and tetrapods [12]. These gave rise to the mammalian class I and II ORs, with humans having 103 class I and 752 class II ORs. Of the total of 855 human ORs, ~370 have intact open reading frame, 425 are pseudogenes and ~60 have both intact and disrupted alleles in the human population (segmenting pseudogenes; HORDE, http://bio-portal.weizmann.ac.il/HORDE/). Based on sequence identity, the mammalian OR superfamily is further classified into families and subfamilies, but the relationships between such classification and odorant specificity remains unclear [70].

Most mammals possess an additional chemosensory apparatus called the vomeronasal organ, with two types of chemoreceptors, V1R and V2R. These are thought to be responsible for the detection of certain pheromones and of non-volatile compounds [6]. In contrast to ORs, V1Rs and V2Rs repertoires vary greatly in size among mammals, with high sequence divergence. For example, the platypus genome contains 270 V1Rs, whereas the dog genome contains only 41. In human, the majority (>95%) of V1R sequences are pseudogenes.

Genetic variation in human olfactory receptors: different noses for different folks

An important corollary of human olfactory repertoire diminution is the evidence that this process is still ongoing. Two types of genomic variation leading to OR inactivation species of trichromatic color vision – highly developed color sensing based on three visual pigments [14]. Nevertheless, the advantage of trichromacy, enabling visual rather than olfactory discrimination of food or mate in higher primates, is still controversial [20].

With fewer active ORs, affinity values for those ORs that react with a given odorant might tend to be more widely spaced (i.e. showing smaller odorant specificity overlaps among repertoire members) [21]. This might result in human threshold variations becoming more prevalent than for mammals with a larger OR repertoire, and a greater likelihood for an inactivating OR mutation or polymorphism to have a discernible phenotype. This potential scenario, together with the facility with which one can measure olfactory capacities in humans (Box 2), should make our species an ideal model system in which to study chemosensory genetics questions.

Box 2. Evaluation of olfactory threshold

Sensory evaluation or psychophysical measurements are the main route for measuring differences in olfactory phenotypes. These include measurements of threshold (the lowest concentration at which an odorant can be detected), magnitude estimation (assessing odor intensity) and measurements of quality perception (type and pleasantness of an odor) [32]. Objective appraisal of threshold usually involves repeated presentation of the odorant to the subject (e.g. in ascending concentrations, seeking a statistically significant discrimination) [71,72]. Olfactory psychophysical measurements are often time consuming owing to the need for time-sparse presentations to avoid adaptation to the odorant.
Box 3. Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs), that is, single base substitutions or insertions or deletions, are the most common type of genomic variation, occurring once in 100–300 bp on average. Currently, ~14.7 million such simple polymorphic events (SNPs and deletion-insertion polymorphisms [DIPs]) are archived in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), and their broader definition includes multi-base DIPs. Only a small fraction of these fall within coding sequences of genes and an even smaller fraction affect protein sequence.

This review highlights the role of SNPs and DIPs that result in an in-frame stop codon, frame-shift mutation or substitution of a highly conserved amino acid, thus leading to severe impairment or complete inactivation (segregating pseudogenes). Such events, which can reach high population frequencies owing to relaxed purifying selection in multi-receptor repertoires, are perfect candidates to underline the variability in olfactory sensitivity and perception among humans. Other terms, such as ‘stop-SNPs’ or ‘X-SNPs’, have been used to describe these variations, which are also found in non-OR genes [22,23]. Combinations of particular SNPs might reveal their phenotypic effects through haplotype (‘haploid genotype’), a physical combination of particular alleles along a chromosome. Usually, the genomic distance between the SNPs is inversely correlated to their tendency to be inherited together without recombination. Such joint passage from one generation to the next (linkage disequilibrium) is relevant to olfactory genetics, as exemplified by the two SNPs in OR7D4, R88W and T133M, which show only two of the four possible haplotypes, RT and WM.

are relevant in this respect: segregating pseudogenes [22,23] and CNVs [24], which involve deletion alleles (Figure 1a and Boxes 3,4). Such variations constitute natural knockout of specific ORs in some individuals, but not in others, potentially leading to phenotypic differences in olfactory acuity and perception. Segregating pseudogenes arise owing to a specific type of single nucleotide polymorphism (SNP) that leads to gene inactivation (Box 3). Approximately 60 OR segregating pseudogenes have been discovered [22,25], including 20 that were previously catalogued as pseudogenes (Box 1). More segregating pseudogenes are expected to be revealed by future deep sequencing efforts (Box 5). An analysis of several hundred human individuals found that almost every person has a unique combination of intact and inactive OR alleles (inactivation ‘barcode’, [22]), which is arguably one of the most pronounced case of genetic variation in humans [26].

The second form of genetic variation, CNVs (Box 4), has recently been reported in several studies addressing the human olfactory subgenome, based on available data from genome-wide surveys and databases [27,28], or derived from experiments with specifically designed high-resolution arrays [29]. These analyses show that ORs are significantly enriched in regions of CNVs. This is either as a result of genomic positional bias towards CNV formation (i.e. ORs tending to reside in genomic regions that have enhanced propensity for CNV generation, irrespective of the ORs themselves), or owing to the relaxation of selection against CNVs. Evolutionary young receptors (i.e. those lacking a one-to-one ortholog in chimpanzee or those with close paralog in the human genome) were found to have more CNVs [29]. This finding might represent one mode of positional bias, whereby an enhanced probability of long-term evolutionary dynamics that involves gene ‘birth and death’ is correlated with a higher propensity of the short-term process of CNV formation in OR-containing genomic segments.

Among the types of CNVs, deletion alleles that affect one or several consecutive intact ORs within a cluster are the most likely to produce a detectable phenotype. Altogether, 12 such alleles that affect 18 intact ORs and five OR pseudogenes were identified [28,29]. The largest deletion, on human chromosome 11, affects six OR genes, of which four are intact. A second locus on chromosome 11 involves two genes, olfactory receptors family 8, subfamily U, members 8 and 9 (OR8U8 and OR8U9) that are fused via a deletion, which is effected by non-allelic homologous recombination within their coding region. This deletion and fusion event creates an allele with a third, apparently intact OR – OR8U1 (Figure 2d). Thus, some individuals can have an allele containing one OR; others will have an allele with two genes, potentially with disparate odorant specificities, both of which are different from those of the fusion product. Consequently, heterozygotes would have three types of sensory neurons bearing these three genes. Gene fusion thus provides a rapid evolutionarily mechanism for the generation of novel paralogs, similar to gene conversion [30], and represents an alternative to the slower pathway of duplication and divergence.

Box 4. Copy number variation

The term CNV usually denotes DNA segments with a length ≥1 kb, present in populations as alleles with variable copy number. Thus, for example, different chromosomes in the human population might harbor zero or two copies of a genomic segment instead of the original one copy. CNVs arise through deletions, duplications, insertions or more complex genomic events. Genome-wide studies indicate that CNVs are responsible for the majority of variable base-pairs among individuals and might be a main basis for phenotypic differences [74], with implications to olfactory phenotypes. It was recently suggested that CNVs have an impact on genome evolution by facilitating the expansion or diminution of gene families [24], thus potentially serving as an additional factor affecting the size of the OR repertoire.

Box 5. Next-generation sequencing

Next-generation sequencing refers to a set of new DNA sequencing technologies that can process millions of sequence reads in one experiment, offering a dramatic ~1000-fold increase in throughput cost-effectiveness relative to the standard methods (for a review, see Ref. [75]). These technologies, producing several gigabases of sequence per one-week run, are expected to increase their capabilities further in the coming few years, fulfilling the goal of a $1000 human genome. As proof of concept, three entire human genomes have been sequenced [76–78] and ambitious projects aiming at complete sequencing of 1000 individual genomes (http://www.1000.genomes.org/) and 50 000 cancer genomes have been launched (http://www.icgc.org/). The wealth of genetic information available from these sequencing efforts begins to highlight the full extent of inter-individual genomic variation. In the realm of genetic association studies, next-generation sequencing brings us to the verge of a new era whereby complete re-sequencing of the human genome, rather than genotyping of a subset of markers, will become feasible.
sensitivity and quality perception. This variability includes differences in general olfactory acuity and in the sensitivity towards particular odorants. The latter was reported as early as a century ago in studies such as the one by Blakeslee [31]. Typically, the distribution of human thresholds (Box 2) towards a particular odorant is bell shaped (Figure 3), whereby one extreme represents diminished sensitivity towards one particular odor (specific anosmia) and the other represents enhanced sensitivity to that odor (specific hyperosmia) [32].

**Specific anosmia – diminished sensitivity towards particular odorants**

Examples of the strongly reduced capacity (10–1000 times) of an individual to detect a particular odorant (specific hyposmia) have been amply documented for dozens of different compounds, the best studied being musk, the sweaty odorant isovaleric acid, and the boar pheromone androstenone. An early study of 1.5 million subjects, perhaps the largest human sensory study ever, provided ample confirmation of this olfactory sensitivity picture [33]. A notable example is the specific anosmia for androstenone (5α-androst-16-en-3-one), with ~30% rates of specific anosmia, in conjunction with pronounced quality perception variation, ranging between extremely unpleasant (urine-like) and pleasant (sweet, floral) [34,35].

Such widespread phenotypic diversity prompted several studies that sought a genetic basis for the threshold and perception variation. An early report showed Mendelian recessive inheritance for pentadecalactone (musk) sensitivity in humans [36]; later, androstenone thresholds were found to be highly concordant in monozygotic twins [37,38] and similar genetic evidence was produced for other odorants [39,40]. For isovaleric acid, one of the best-studied odorants, association was found between sensory thresholds and variability at two distinct genomic loci, on mouse chromosomes 4 and 6. In parallel, environmental and behavioral factors have been suggested to affect olfactory performance, in line with a complex trait [41–44]. Reports that some subjects, ostensibly anosmic to androstenone, became capable of perceiving it after repeated exposure [35] supports the notion of complex trait variation at the sensory receptor level, and contributed to making androstenone an appealing candidate for extensive search for the genetic determinates of odor sensitivity and perception.

A novel approach [45] employing a combination of *in vitro* experiments with genetic association studies convincingly showed that a combination of two non-synonymous SNPs (R88W and T133M) in the human OR gene *OR7D4* (Figure 1a) accounts for 19–39% of the variation in sensitivity and quality perception of androstenone. The study found that subjects with at least one copy of the WM haplotype (Box 3) are less sensitive to androstenone than those that do not carry a WM allele, and the former also report the odorant to be less unpleasant.

These results strongly indicate that *OR7D4* is the highest affinity receptor for androstenone within the human OR repertoire, and provide for the first time the link between genetic variation in OR and odor perception, both *in vivo* and *in vitro*. Although these SNPs are not in the putative binding site, *in vitro* assays indicate that each of the SNPs affects *OR7D4* function [45], and that the *in vivo* phenotype is the result of their combinatorial effect. In addition, the study showed that other amino acid substitutions in the extracellular loop 2 might influence *OR7D4* function *in vitro*, indicating that these residues might be involved in androstenone interaction with the receptor.

The sweaty odorant isovaleric acid is one of the first to come up with a clear-cut evidence for specific anosmia in humans [46]. A recent genetic study [47] showed an association signal between isovaleric acid sensitivity and the genotype of a segregating OR pseudogene *OR11H7P* on...
human chromosome 14. The intact form of OR11H7P, heterologously expressed in amphibian cells, showed preferential response to isovaleric acid in vitro assays. However, the human genomic interval containing OR11H7P is not syntenic to any of the intervals reported to be linked to isovaleric acid sensitivity in the mouse [48]. Intriguingly, the human association signal seems to arise from the paucity of the OR11H7P disrupted homozygotes in the hyperosmic group, one of the first genetic results addressing olfactory hypersensitivity. In general, it is conceivable that specific anosmia and hyperosmia are two sides of the same coin: for specific anosmia, a highest affinity OR can be mutated in a minority of the population, whereas in specific hyperosmia, such receptor might be intact only in rare cases.

**General anosmia**

In addition to differences between individuals in their ability to perceive specific odorants, humans also vary in their general olfactory capabilities. This covers a wide range of phenomena, from general anosmia and hyposmia through to general hyperosmia [49–53].

More than 1% of the Western world population has some form of olfactory disorder. Most of the patients have an acquired condition, which might develop owing to non-genetic causes, such as allergy, viral infection, sinus diseases, head trauma or noxious chemical effects [52,53]. A much smaller proportion has congenital general anosmia (CGA), an inherited loss of smell, which might occur syndromically (i.e. in conjunction with other anomalies e.g. Kallmann’s Syndrome [54]), or appear as an isolated condition. The latter type has a population frequency of ~1:10 000 and often seems to be associated with olfactory epithelial degeneration [55,56]. CGA results in serious deficits in the enjoyment of fragrance, food and beverage, as well as in detection of risk of fire and poisoning. Genetic studies of isolated CGA are very limited owing to its low prevalence, and although no causative mutations were found for this phenotype, all familial cases described are consistent with an autosomal dominant mode of inheritance with incomplete penetrance [57,58].

In mouse, inactivation of any of three transduction genes in the olfactory GPCR pathway, namely, the cyclic nucleotide-gated channel (Cnga2) [59], the stimulatory olfactory G-protein (Gnal or Golf) [60], and adenyl cyclase III (Adcy3) [61] display profound reductions or even absence of physiological responses to odorants. A study of isolated CGA families [57] examined the coding regions...
of the three orthologous human olfactory transduction genes CNGA2, ADCY3 and GNAL but did not find any mutations. Thus, in humans, CGA might be chiefly caused either by mutations in the non-coding regions of these genes, or by mutations in other genes that are essential for proper formation and function of olfactory sensory neurons.

**General olfactory threshold variation**

When testing several unrelated odorants in numerous individuals, a significant inter-odorant threshold correlation has been reported [47,62]. Thus, a lower threshold (higher sensitivity) to one odorant in an individual predicts a tendency of the same individual to have lower thresholds for other odorants (Figure 3). Such observation implies the existence of a shared mechanism that affects an individual’s overall olfactory sensitivity. This is consistent with the proposal [62] that a general olfactory factor can modulate the outcome of odorant-specific threshold measurements. A general sensitivity-modulating mechanism might explain some of the anecdotal evidence for considerable inter-individual variation in overall olfactory performance. Although such variability could be partly related to non-genetic mechanisms, an appealing possibility is the contribution of genetic polymorphisms (in the signal transduction genes, for example, as well as in genes involved in propagation and processing of the olfactory input, such as those underlying the development of olfactory epithelium and olfactory bulb [57]). A similar mechanism has been shown to lead to overall reduction of mouse auditory sensitivity owing to targeted deletion of the prestin gene (Slc26a5), which results in the loss of outer hair cell electromotility [63].

**Concluding remarks**

Odorant-specific sensitivity variations and, in all likelihood, general sensitivity differences are highly prevalent, whereas congenital general anosmia is rather rare [42,64]. Although some efforts will probably be directed towards understanding the diverse molecular mechanisms that might underlie the general deficits, we predict that most research will focus on odorant-specific sensitivity phenotypes, with their obvious causative molecular target – OR genes. Thus, future studies are likely to bring forth a much better understanding of genotype–phenotype relationships in the OR universe.

Olfactory attributes are polygenic in nature, as exemplified by the finding that heterozygotes for the OR7D4 WM allele show intermediate sensory measure values. Nevertheless, future studies will probably lead to breakthrough understanding, employing both monogenic and polygenic models. Furthermore, future extension of studies of ethnic differences in the prevalence of intact ORs compared with pseudogenes [22] could reveal ethno-geographic facets of human olfaction.

OR-odorant relationships can be revealed either through genetic association [45,47] or by in vitro heterologous expression [65,66]. In the former, subject recruitment and psychophysical testing are time-limiting because a rather large sample size (several hundred) is required for adequate statistical power. As a result, only a few odorants are typically tested and the probability of a positive genetic association is rather low. However, high-throughput identification of specific OR-odorant pairs has recently become feasible because of important improvements in heterologous expression technologies [45,65–67], enabling the testing of >300 ORs with several dozen odorants. Future studies should integrate both approaches by focusing on pre-screened odorants for receptors with high genetic variability.

It will soon become feasible to discover or score practically all the OR-related genomic variation in exons, introns and transcription control regions using cutting edge methodologies (Box 5). This combined progress will enable a deeper understanding of the odor space and genomic variation, helping to elucidate the mysteries of human olfaction.

**Acknowledgements**

We would like to thank B. Brumshtein for the illustration of OR7D4 structure (Figure 1), and Edna Ben-Asher for critical reading and remarks. Supported by NIH (NIDCD) and the Crown Human Genome Center at the Weizmann Institute.

**References**