Ecology and Evolutionary Biology
2021; 6(3): 53-77
http://www.sciencepublishinggroup.com/j/eeb
doi: 10.11648/j.eeb.20210603.11
ISSN: 2575-3789 (Print); ISSN: 2575-3762 (Online)

An Evolutionary Based Strategy for Predicting Rational Mutations in G Protein-Coupled Receptors

Miguel Angel Fuertes*, Carlos Alonso

Department of Microbiology, Centre for Molecular Biology "Severo Ochoa", Spanish National Research Council and Autonomous University, Madrid, Spain

Email address:
calonso@cbm.csic.es (C. Alonso), mafuertes@cbm.csic.es (M. A. Fuertes)

*Corresponding author

To cite this article:
Miguel Angel Fuertes, Carlos Alonso. An Evolutionary Based Strategy for Predicting Rational Mutations in G Protein-Coupled Receptors. *Ecology and Evolutionary Biology.* Vol. 6, No. 3, 2021, pp. 53-77. doi: 10.11648/j.eeb.20210603.11

Received: April 24, 2021; **Accepted**: May 11, 2021; **Published**: July 13, 2021

Abstract: Capturing conserved patterns in genes and proteins is important for inferring phenotype prediction and evolutionary analysis. The study is focused on the conserved patterns of the G protein-coupled receptors, an important superfamily of receptors. Olfactory receptors represent more than 2% of our genome and constitute the largest family of G protein-coupled receptors, a key class of drug targets. As no crystallographic structures are available, mechanistic studies rely on the use of molecular dynamic modelling combined with site-directed mutagenesis data. In this paper, we hypothesized that human-mouse orthologs coding for G protein-coupled receptors maintain, at speciation events, shared compositional structures independent, to some extent, of their percent identity as reveals a method based in the categorization of nucleotide triplets by their gross composition. The data support the consistency of the hypothesis, showing in ortholog G protein-coupled receptors the presence of emergent shared compositional structures preserved at speciation events. An extreme bias in synonymous codon usage is observed in both the conserved and non-conserved regions of many of these receptors that could be potentially relevant in codon optimization studies. The analysis of their compositional structures would help to design new evolutionary based strategies for rational mutations that would aid to understand the characteristics of individual G protein-coupled receptors, their subfamilies and the role in many physiological and pathological processes, supplying new possible drug targets.

Keywords: Evolution, Triplet Composon, Ortholog, G-protein Coupled Receptor, Odorant Receptor

1. Introduction

G protein-coupled receptors (GPCRs) are integral membrane proteins used by cells to convert extracellular signals into intracellular responses. Inferred from their sequence and their structural similarities these receptors form a superfamily of membrane proteins containing many different families [1, 2] amongst which the aminergic GPCRs, are targets for ~25% of current drugs [3]. The analysis of sequences in an evolutionary context suggests that dimerization appears to occur quite generally in GPCRs [4]. More than 50% of GPCRs are orphan receptors [5, 6] sharing common structural motifs in which seven transmembrane helices are connected to three extracellular loops and to three intracellular loops (see Figure 1A), being activated by a variety of external stimuli [7]. In general, they are targets for the bulk of therapeutic drugs, playing important roles in many biological processes [8, 9].

GPCRs tend to have conserved functional motifs in mammalian genomes making them distinguishable from other proteins. They present conserved compositional patterns in transmembrane helices and in extracellular and cytosolic loops [10]. Even though many members of the GPCR family have a large number of functional and structural similarities, they lack extensive sequence similarities. Thus, while over 40 GPCRs have been crystalized, the structure of not even one member of the largest GPCRs family, the olfactory receptors (ORs), has been determined. The poor functional heterologous expression, the post-translational modifications, the high flexibility, and the low stability in detergent hampering the understanding of the molecular mechanism of the OR

function [11-13]. The ability to release and identify specific smells is a key feature for survival of most animals [14]. The mechanism of odor recognition is still lagging and will require solving multiple structures in complex with native ligands. In fact, the residues involved in the dynamic process that converts an inactive OR structure into an active one are still obscure. The latest advances in strategies leading to the structural resolution of GPCRs makes this goal to be realistic [15]. The high-resolution structures of specific GPCRs have provided insights that help the understanding of the relationship between the chemical structure of a ligand and the way it affects the conformation of the receptor [16]. However, there are not available crystallographic structures of ORs, therefore, most studies regarding the mechanistic analysis of ORs function have to be relied on the use of molecular dynamic modelling [17]. Certain clues regarding potential residues involved in the OR activation have been tentatively proposed, but they remain to be assessed by means of in vitro experiments and molecular dynamic simulations [18-19].

Recently, an analysis of human-mouse orthologs by the triplet composon method (or tCP-method) has shown new categorizations characterized by specific tCP-usages [20]. The bulk of human-mouse orthologs analyzed were categorized in 18 clusters [21]. One of those categorizations contains mostly genes coding for transmembrane proteins, in their majority GPCRs, and mainly ORs [21] although a wide-range categorization of these receptor have not carry out yet. Lately, a variant of the tCP-method has been used to search for common DNA patterns distributed along the length of orthologs. In a large number of cases, appearing shared nucleotide structures between orthologs interspersed along the protein coding ORFs [22, 23]. The aim of the study was to identify novel compositional structures interspersed in GPCRs that could be useful for site-directed mutagenesis and potentially relevant in seeking new drug targets.

2. Methods

2.1. Gene Sample

GPCRs were downloaded from the HUGO Gene Nomenclature Committee (HGNC) https://www.genenames.org/gene families/GPCRA#ADRAA at the European Bioinformatics Institute (EBI) [24]; additional entries were downloaded from the Human Olfactory Data Explorer (HORDE) http://bioportal.weizmann.ac.il/HORDE// [25], the GeneCards database from the Weizmann Institute of Science [26], with Copyright ©1996-2018, http://www.genecards.org/, and the Ensembl release 92, April 2018 ©EMBL-EBI, http://www.ensembl.org. We only include in our study DNA from the longest transcripts.

The dataset used in the study contain 546 and 573 human and mouse GPCRs, respectively, among which 200 are human ORs and 400 mouse ORs (Table 2). The orthologs are identified by the name, the protein description and the number of shared tCPs. The table shows many human ORs having several mouse orthologs, and vice versa [27]. The parameters to state orthology in the Ensembl databank are in http://www.ensembl.org. Working with ortholog genes are needed among species for drawing genomic changes along their evolution. The comparison of mouse and human orthologs is of particular interest because mouse is often used as a model organism to understand human biology and however, both species are evolutionary distant [28]. The nomenclature used is the commonly used for human olfactory receptors [29], e.g. OR3A1 would correspond to the human olfactory receptor, family 3, subfamily A, member 1. In mouse we use, to avoid confusion, the official symbols given by the Mouse Genome Informatics (MGI) Nomenclature Group from the Jackson Laboratory, Bar Harbor, Maine, http://informatics. jax.org [30], e.g. *Olfr10*.

Table 1. *tCP code. List of composons (tCPs) and their associated nucleotide triplets.*

tCP	DNA-triplets associated to tCPs
<A>	AAA
<T>	TTT
<G>	GGG
<C>	CCC
<AC>	AAC, CAA, ACA, CCA, ACC, CAC
<AT>	AAT, TAA, ATA, TTA, ATT, TAT
<AG>	AAG, GAA, AGA, GGA, AGG, GAG
<CG>	CCG, GCC, CGC, GGC, CGG, GCG
<GT>	GGT, TGG, GTG, TTG, GTT, TGT
<CT>	CCT, TCC, CTC, TTC, CTT, TCT
<AGC>	AGC, GCA, CAG, ACG, CGA, GAC
<AGT>	AGT, GTA, TAG, ATG, TGA, GAT
<ACT>	ACT, CTA, TAC, ATC, TCA, CAT
<TCG>	TCG, CGT, GTC, TGC, GCT, CTG

2.2. Brief Description to the tCP Method

The tCP method counts NT triplets of a genomic sequence read fully overlapping categorizing them by their gross composition (Table 1). The justification for this method is theoretical and based on the existence of exclusionary multiplets characterized by the presence or absence of particular bases. The number of categorizations (called 'triplet-composons' or tCPs), resulted to be 14 and the optimal length for the multiplet was 3 NTs (a triplet) [20]. In summary, 14 tCPs contain all possible nearest-neighbors that can be formed with the four DNA bases. In order to take into account all and every one of the contexts of each NT, the sequence is read 'fully overlapping', guaranteeing that all triplets be considered to avoid losses of information. We assume, when reading the sequence in non-overlapping way, that the lost triplets could supply relevant evolutionary information. The notation of tCPs and their associated sets of NT-triplets are in Table 1.

2.3. Numerical Analysis

As an example, the distribution of tCPs in the human Beta-adrenergic receptor (β_2AR) codified by the ADRB2 gene, was calculated (Figure 1a). The receptor mediates the catecholamine-induced activation of adenylate cyclase through the action of GPCRs [31-33]. We have used a variant

of the original tCP-method [21] to analyze the cumulative tCP-usage frequency distribution (Figure 1B) along the gene sequence [22, 23]. In the figure it can be observed the cumulative graph and the trend line for the usage frequency of the tCP <AG>. The cumulative tCP-usage would be, then, the sum of all previous tCP-appearances along the sequence up to the current length. In order to facilitate the study of the distribution of tCPs along of the ADRB2 sequence, the cumulative tCP-graph and the trend line (Figure 1b) were projected on the length axis (Figure 1c) by subtracting the cumulative tCP-usage frequency from the regression line, hereafter named 'tCP-profile' [22-23]. The tCP-profile represents the distribution of differences between the tCP-events observed, *tCPo*, and the estimated, *tCPe*. We will consider that two ortholog sequences share a similar tCP-profile when their Pearson correlation coefficient (r) is equal or higher than $r \geq 0.85$. Coincidences in short stretches of the tCP-profile reduce notably the correlation coefficient of the complete sequence. We sought to impose a sufficiently stringent cut-off because it is known that an excess similarity implies common ancestry and consequently two homologous sequences showing a high correlation coefficient would guarantee a causal link with function [34]. On the other hand, in our GPCR sample, we want to avoid many ortholog sequences with low similarity but clearly homologous based on statistically significant structural similarity or strong sequence similarity to an intermediate sequence. Graphs and statistics were carried out with the package OriginPro 8 SRO V8.0724 (B724) ©OriginLab Corporation. Each one of the 14 tCPs listed in Table 1 has its own characteristic tCP-distribution that allows us to compare them between species. We represent in a panel the comparison of the 14 different tCP-profiles of each human-mouse ortholog from the dataset.

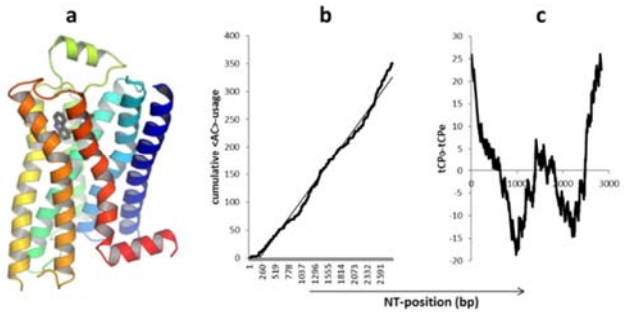

Figure 1. *(a) The human beta-2 adrenergic receptor (ADRB2) in complex with the partial inverse agonist carazolol [31-33]. (b) Cumulative plot of events for the tCP <AG> in ADRB2 (thick line) and their corresponding trend line (thin line). (c) Profile of ADRB2 for <AG> representing the distribution of differences between the tCP-events observed, tCPo, and the estimated, tCPe.*

2.4. Differences Between the tCP Method and Current Methods

The tCP method deeply differs from other current methods, not in the sequence comparison metrics but what is the object of measure. Usually, we measure identity between sequences when the object of the measurement is, for example, a NT sequence. In the tCP method the sequence comparison metrics does not change at all but does so the measure object; it is not a NT sequence but a tCP sequence. Thus, the first task to be done would be to transform a NT sequence to a tCP sequence. To do that, it is necessary read ortholog sequences in "fully overlapping way" stablishing for each nucleotide all their nearest neighbors and categorizing them by their gross composition by using Table 1. We count, then, the frequency of appearance of each tCP along both sequences to obtain from the profiles, the conserved tCPs; this let us to locate the conserved stretches in both tCP sequences. Afterwards, we locate the conserved fragments in the NT sequences.

2.5. Criteria Used to Highlight the Shared tCPs Along the Gene Length

The tCPs shared by orthologs were obtained as follows: i) NT-sequences of human-mouse orthologs were translated into tCP-sequences using Table 1. ii) To compare sequences, we use the dynamic algorithm shaped for the global alignment of two sequences [35, 36]. iii) A graphical representation of the data (the tCP-profile) was made, and iv) the level of similarity between the aligned tCP-sequences of each human-mouse ortholog was computed using the Pearson correlation coefficient (r). As control of non-random tCP-distributions along the genes, we analyzed the tCP-profiles of randomly generated sequences having NT-compositions identical to the orthologs. The utility *shuffleseq* from the EMBOSS explorer was used to generate the random sequences maintaining the NT composition [37, 38]. We confirmed also that identical NT-sequences have identical tCP-profiles as expected (data not shown).

3. Results

3.1. tCP-Compositional Patterns in a Representative Ortholog GPCR

To determine whether there are tCP-compositional patterns shared by human-mouse ortholog-GPCRs requires to generate the tCP-profile of each pair of ortholog sequences in order to be compared. To illustrate the process, we will use, as an example, the human-mouse ortholog, ADRB2 (Figure 1a). This ortholog shows high NT-percent similarity (85%), moderate tCP-percent similarity (67%) and similar tCP profiles in both species (see Table 2).

Figure 2 shows a panel representing the 14 tCP-profiles of ADRB2. The numerical inset shows the correlation coefficients of each tCP when comparing both orthologs; 3 out of 14 tCPs show correlation coefficients higher than the cut-off, namely, <AG>, <GT> and <AGC> showing similar profiles to each other; consequently, many of NT-triplets associated with those tCPs would be shared by both orthologs. The remaining tCPs show correlation coefficients ranging from r=0.36 for <T> to r=0.83 for <CT>. Although only tCP-profiles showing correlation coefficients higher than the cut-off were considered, we also observed short

highly correlated fragments in some tCP-profiles, as it would be the case of <CT>, <TCG> and <GC>, with correlation coefficients near but lower the cut-off; they were discarded from the analysis although they contribute substantially to the percent of nucleotide identity between both orthologs. Lower correlation coefficients ranging from 0.36<r<0.63, were also observed for tCPs <A>, <T>, <G>, <C>, <AC>, <AT> and <ACT>.

We have obtained in a similar way to ADRB2, the shared tCP-profiles of all human-mouse pairs of ortholog sequences from the dataset (see Table 2).

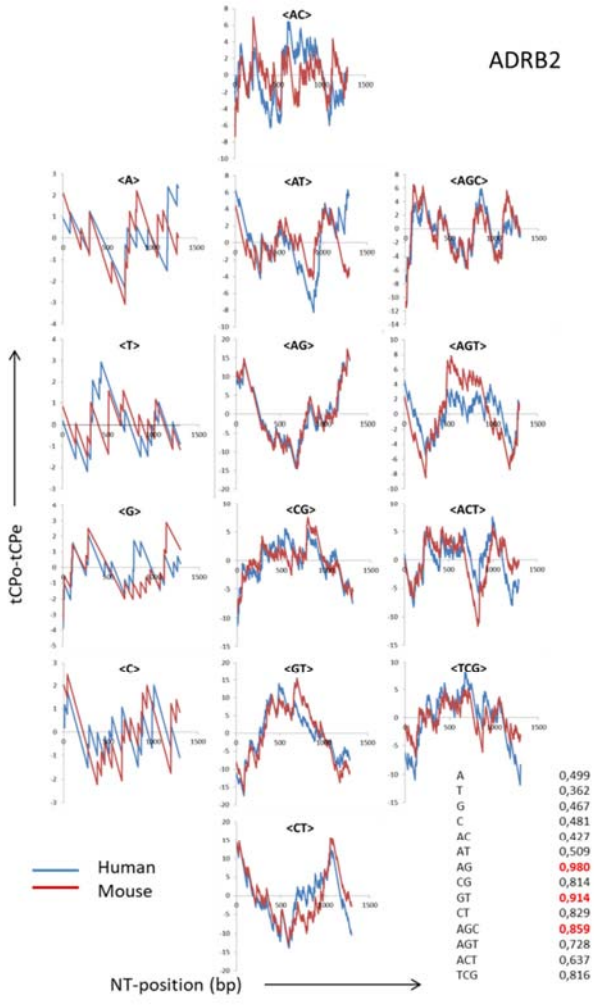

Figure 2. *Comparison of tCP-profiles from the human-mouse ortholog ADRB2. Red and blue lines correspond to the profiles of mouse and human, respectively. The insets display the name of the ortholog (upper right corner), and a table with the Pearson correlation coefficient, r, between tCP-profiles of human and mouse (lower right corner). Pearson correlation coefficient for r≥0.85 are in red.*

3.2. The Conserved tCP Structure Between Orthologs is Under Selection Constraints in GPCRs

We investigated the emergent compositional structures shared by human-mouse ortholog GPCRs to study whether there are evidences of constraints for variability. NT-mismatches in three human-mouse orthologs of the LGR family were investigated, and also three genes of the OR family, namely, LGR4, LGR5, LGR6, OR52E6 (*Olfr671*), OR8J (*Olfr1045*) and OR7A10 (*Olfr1351*), all of them different in their tCP usage and in the type and number of conserved tCPs. It was observed that LGR4 conserves 2 tCPs, namely, <G> and <TCG>, LGR5 conserves 8 tCPs, <A>, <T>, <AC>, <AT>, <AG>, <CG>, <GT>, and <CT>, LGR6 conserves 8 tCPs, <A>, <AC>, <AT>, <AG>, <CG>, <GT>, <ACT> and <TCG>, OR52E6 (*Olfr671*) conserves 2 tCPs, <T> and <CT>, OR8J1 (*Olfr1045*) conserves 4 tCPs, <G>, <AT>, <AG> and <AGT> and finally, OR7A10 (*Olfr1351*) conserves 2 tCPs, <G> and <AC>. To see the tCPs conserved in more ortholog GPCRs see Table 2. The analysis compares the percent of NT mismatches in regions in which the orthologs analyzed share the conserved tCPs, relative to the percent of NT mismatches observed in those regions in which only one of the genes contains the conserved tCPs. The results are in the bar diagram of Figure 3 and show that in GPCRs, the ortholog shared regions present low percentages of NT mismatches, in contrast with those regions in which only one ortholog have the conserved tCPs being the difference highly significant in average ($p<0.0001$). Although evolutionary changes occur at speciation events, data suggest, however, that shared regions would be under evolutionary constrains for variability reinforcing the idea that such compositional structures would be conserved at speciation events.

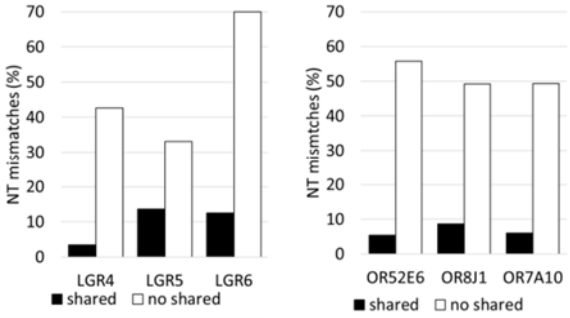

Figure 3. *Bar diagram showing the percent NT mismatches in regions in which the orthologs compared share the conserved tCPs (black bars) relative to the percent NT mismatches in regions in which only one of the orthologs contains conserved tCPs (white bars).*

3.3. Statistical Data of Shared tCPs from Human-Mouse Ortholog-GPCRs

Figure 4 shows the statistical data obtained from the Table 2. When representing the percent of GPCRs sharing a given tCP (Figure 4a) we observed that <AG> is the tCP shared for more than 58% of human-mouse orthologs, indicating that, at speciation events, although evolutionary changes occur more frequently, there are in GPCRs long gene regions evolutionary constricted to maintain the use of triplets containing all nearest neighbors of A and G. Apart from the tCP <AG>, tCPs more frequently shared are <CT> (35%) and <GT> (27%) respectively. On the other hand, <AGC> (12%) and <AT> (13%) respectively, are the less shared tCPs.

Figure 4b shows the percent of GPCRs sharing a given

number of tCPs. The study of ORs and GPCR-non-ORs shows significant differences in the percent of shared tCPs. Thus, 75% ORs share less than 4 tCPs while most of all ORs do not share more than 5 tCPs. However, 70% GPCR-non-ORs share less than 6 tCPs while practically all of them do not share more than 9 tCPs. None of the OR shares more than 8 tCPs in contrast with GPCR-non-ORs that share up to 12 tCPs as is the case of FZD4, a gene codifying for the frizzled-4 receptor (see Table 2). It may be due to the restrictive cut-off imposed. Only a few ORs (12%) and GPCR-non-ORs (6%) do not share any tCP for the cut-off considered.

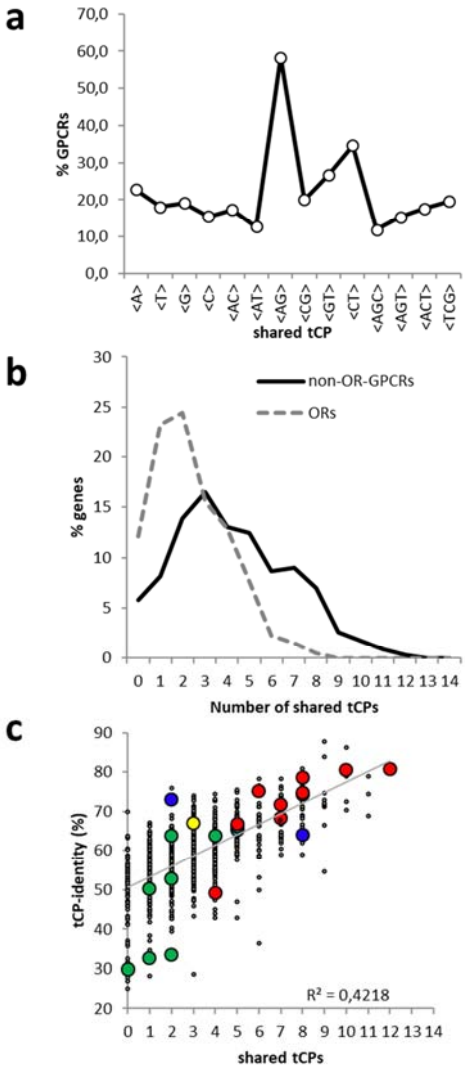

Figure 4. *Information collected from the ortholog dataset. (a) Plot representing the percent orthologs sharing a given tCP. (b) Graph representing the percent orthologs sharing a given number of tCPs. The dashed line (----) represents ORs and the continuous line (——) represents GPCR-non-ORs. (c) Point plot representing, for each ortholog, the number of tCPs shared between species against their percent of tCP similarity. The grey line represents the regression line and the corresponding determination coefficient R^2. Colored big dots represent specific orthologs considered in more detail: LGR4 and LGR5 (dark blue), mouse orthologs of the human OR3A1 (green), frizzled orthologs (red), human ADRB2 (yellow).*

Figure 4c shows a point graph representing the percent similarity of each sequence alignment against the number of shared tCPs. As it can be observed, for a given number of shared tCPs there are a notable number of ortholog-GPCRs with different percent similarity. For example, for a percent similarity of 60% we can find ortholog-GPCRs sharing a number of tCPs ranging from 0 to 9. The notable dispersion observed suggests that there is a significant degree of independence between the percent similarity and the amount of shared tCPs. In any case, the regression line shows a positive slope indicating that in some cases higher identities correlate with higher number of shared tCPs as it would be expected. However, this is not the general rule. The dispersion in the percent similarity is higher as the number of shared tCPs decreases because a higher number of ortholog-GPCRs share a lower number of tCPs as it was mentioned in Figure 4b.

An analysis of Figure 4c and Table 2 let us to obtain evolutionary data from functionally similar ortholog GPCRs having high percent NT-identity and however very different number of shared tCPs. The amount of ortholog sequences meeting those requirements is not very extensive (Table 2). We selected two human GPCRs, LGR4, LGR5 and their mouse orthologs, to carry out a more detailed study. We have not special reasons for the election of LGR4 and LGR5 beyond illustrate an analysis of a pair of related orthologs with very dissimilar compositional characteristics (Figure 4c) and because they contain a large leucine-rich extracellular domain for hormone or neurotransmitters binding [39, 40], attributes making them suitable for a more detailed analysis about the synonymous codon usages of leucine in the different domains of the protein.

3.4. Ortholog-GPCRs Highly NT Identical: The Case of the Leucine Rich Repeat Containing GPCRs LGR4 and LGR5

We will study in LGR4 and LGR5 and their mouse orthologs: i) the dissimilarities among them as well as the number of shared tCPs; ii) the evolutionary constrains observed in the shared regions and iii) the synonymous codon usages in the tCP conserved regions. LGR4 and its mouse ortholog are identical in length (2856 base pairs) and also LGR5 and its ortholog (2724 base pairs); they are highly similar in NT-sequence (89% between LGR4 and its ortholog and 84% between LGR5 and its ortholog). Figure 5a shows a large extracellular domain in LGR4 and LGR5 with multiple leucine-rich repeats mediating ligand interaction, in addition to seven transmembrane domains codified in exon 18 of both genes, and an intracellular domain for signal transduction [41]. The main dissimilarities between LGR4 and LGR5 (Figures 5b and 5c) have to do with the number of shared tCPs; two tCPs shared by the human-mouse ortholog LGR4 (<G> and <TCG>) and eight tCPs by the human-mouse ortholog LGR5 (<A>, <T>, <AC>, <AT>, <AG>, <CG>, <GT> and <CT>). As can be observed, LGR4 and LGR5 do not share any tCP for r≥0.85 despite the fact that both are adrenergic receptors, both are part of the same evolutionary family and they are similar in 3D structure. Figures 8 and 9

illustrate the complete panels showing the shared tCPs of LGR4 and LGR5 and their mouse orthologs. The high correlation observed in the shared tCPs is consequence of the full length comparison of both LGR4 and LGR5 and their mouse counterparts.

3.5. Synonymous Codon Preferences for Leucine in the Human-Mouse Orthologs, LGR4 and LGR5

LGR proteins harbor a distinctive large extracellular domain with 17 leucine-rich repeats composed of 24 amino acids [42]. We search for possible similarities and dissimilarities in the number and the type of synonymous codons encoding leucine in this external region relative to the transmembrane domain and/or the inner domain of binding to G-protein (Figure 5a).

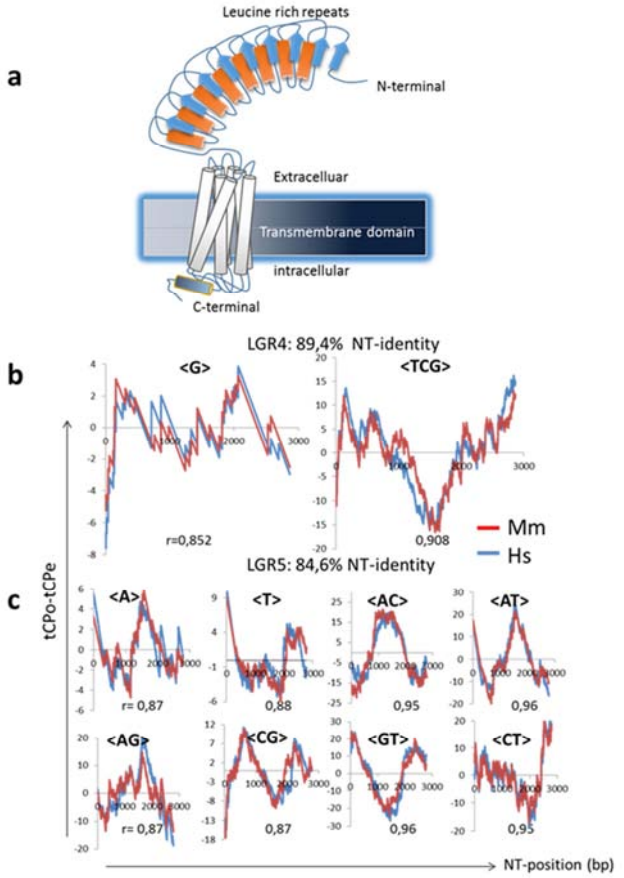

Figure 5. *(a) Scheme showing the structure of the LGR family. Shared tCPs in LGR4 (b) and LGR5 (c); red and blue lines represent tCP-profiles of the mouse and the human, respectively. Identification of shared tCPs and their Pearson correlation coefficients between tCP-profiles (upper and lower side of each profile, respectively). Percent NT-similarity (in upper part of graphs (b) and (c))*

Figure 6 shows the differences observed in the synonymous codon-usage for leucine, in human LGR4 and LGR5, inside and outside the regions containing the conserved tCPs in both orthologs (Figures 4b and 4c). LGR4 conserves the tCPs <G> and <TCG>. It can be observed that <G> and <TCG> contain a unique codon codifying for leucine, the CTG codon, (Table 1). Consequently, the CTG codon would be always 100% preferred inside the regions containing the 2 conserved tCPs <G> and <TCG> being absent outside these regions. The remaining synonymous codons encoding leucine TTA (23%), TTG (25%), CTT (20%), CTC (13%), and CTA (19%) appear exclusively outside the regions containing the conserved tCPs of LGR4, representing a total of 100%, as expected. On the other hand, LGR5 conserved the tCPs <A>, <T>, <AC>, <AT>, <AG>, <CG>, <GT>, and <CT>. It can be observed that those tCPs contain four codons codifying for leucine, CTC, TTG, CTT and TTA, (Table 1). Consequently, those codons would be always preferred inside the regions containing the conserved tCPs with percentages CTC (30%), TTG (25%), CTT (26%) and TTA (19%) representing a total of 100% being absent outside these regions. The remaining synonymous codons encoding leucine in LGR5, CTG (83%) and CTA (17%) appear exclusively outside the conserved regions, representing also a total of 100%.

Results as those obtained for LGR4 and LGR5, are obtained for other human-mouse orthologs with multiple leucine-rich repeats sharing some tCP-profiles. The extreme synonymous codon preference for leucine occurs, then, not only in the external region of the multiple leucine-rich repeats but in all domains of these receptors being a consequence of the tCPs categorization rule (Table 1). The result can be generalized to any ortholog GPCR sharing some tCP-profiles.

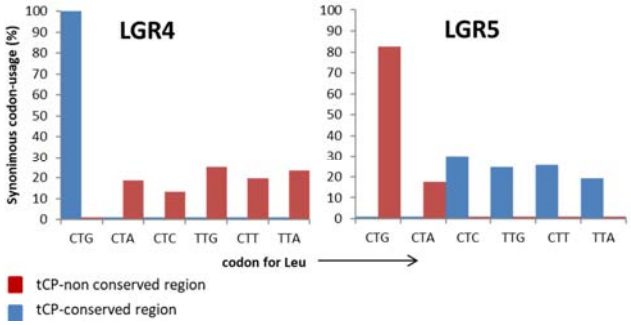

Figure 6. *Differences in the synonymous codon-usage for leucine in the tCP conserved (blue bars) and non-conserved (red bars) regions of LGR4 and LGR5.*

3.6. The issue of Identifying Human-Mouse OR Orthologs by Automated Methods

Many orthologs used in the study would be inaccurate because databases work, in many cases, with automated methods for ortholog identification. The decision (about to be or not to be an ortholog) depends on the parameters used by those automated software's, given sometimes inaccurate results to each other. Ensembl and GeneCards databases were used in this paper and we observed that, in fact, such dissimilarities exist. We think, maybe the research carried out in this paper could shed light in this area. Table 2 shows that a given human OR could have one or more ortholog in mouse and vice versa [27]. However, in different databanks there are dissimilarities about the amount of orthologs between both species. As the decision depends on the

parameters used by those automated methods, we will try to study the behavior of the tCP method in this cases.

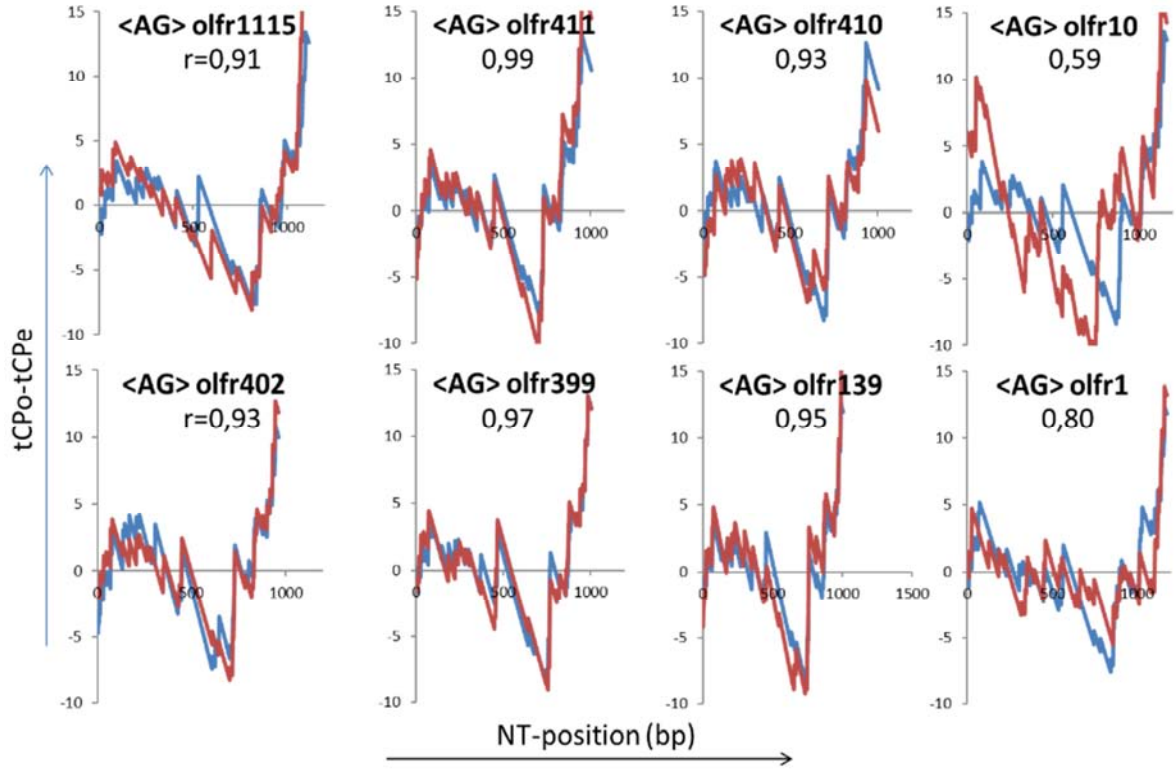

Figure 7. *Red and blue lines correspond to tCP-profiles of the human OR3A1 and their mouse orthologs Olfr1115, Olfr411, Olfr410, Olfr402, Olfr399, Olfr139 and Olfr1, respectively. The upper side of each tCP-profile shows the identification of the shared tCP, <AG>, the name of the mouse ortholog and the Pearson correlation coefficients (r) between OR3A1 and its mouse orthologs.*

As an example, GeneCards indicates that mouse *Olfr1115, Olfr411, Olfr410. Olfr10. Olfr402, Olfr399, Olfr139* and *Olfr1* would be orthologs of the human OR3A1, a gene included in the human GPCR-1 family. On the other hand, Ensembl databank indicates that only *Olfr402* and *Olfr410* would be orthologs of OR3A1. A tCP-profile analysis of those genes shows, however, that the tCP <AG> would be common to *Olfr1115, Olfr411, Olfr410, Olfr402, Olfr399* and *Olfr139* with r≥0.85 but <AG> would be common neither to *Olfr10* nor *Olfr1*; both genes show correlations lower than the cut-off when they are compared with OR3A1 (Figure 7). In fact, *Olfr10* does not have any shared tCP with OR3A1, and *Olfr1* shows two shared tCPs, <AT> and <CG>, but none in common with OR3A1 (Table 2). High correlation coefficients were obtained when <AG>-profiles of OR3A1 and those of *Olfr1115, Olfr411, Olfr410. Olfr402, Olfr399* and *Olfr139* were compared; in all cases r≥0.91 except *Olfr1* (r=0.80) and *Olfr10* (r=0.59), respectively. Looking at the *Olfr1* profile we observe certain resemblance with OR3A1 (r=0.80) despite the fact that, at speciation events, the human gene changed significantly relative its mouse ortholog exhibiting significant variations in composition. In the case of *Olfr10*, we think that changes accumulated are so significant (r=0.59) that we have doubts about whether OR3A1 and *Olfr10* can be considered orthologs. Based on the high correlation observed, we would conclude that *Olfr1115, Olfr411, Olfr410. Olfr402, Olfr399* and *Olfr139* could be orthologs of the human OR3A1 because two orthologous sequences showing a high correlation coefficient would imply a common ancestry [34]; *Olfr10* and *Olfr1*, however, would be dubious, although in this case could not be discounted that they be orthologs since it assumes that two sequences with low similarity can be orthologous based on statistically significant structural similarity or strong sequence similarity to an intermediate sequence [34].

4. Discussion

The amount and importance of GPCRs have interested of experts of many fields [15, 43, 44]. Efforts have been made to elucidate the structure of GPCRs [45]. However, the low number of crystalized GPCRs and the absence of crystallographic data in ORs requires that most mechanistic studies rely on the use of molecular dynamic modelling combined with site-directed mutagenesis data [43] despite the number of restrictions that this type of studies convey. Limited attention has been devoted in the development of new evolutionary based strategies for predicting rational mutations in GPCRs. This is a critical aspect for the interpretation of pharmacological and physiological experiments if conclusions are taken between species [46]. To this end, we will discuss an evolutionary strategy that could be useful to predict rational mutations in GPCRs in general, and in ORs in particular.

The strategy consists in considering all sets of NT triplets with identical gross composition (Table 1); then, reading the

sequence fully overlapping, we count the cumulative frequency of appearance of each one of the sets along the sequence making a comparative graphic (the tCP-profile) between the analyzed sequence and its ortholog (see Materials and Methods). Figure 2 illustrates the tCP-profile of the GPCR ADRB2 and its mouse counterpart, codifying for the β2AR adrenergic receptor [47]. Only three tCPs show correlation coefficients higher than the cut-off and only one, the tCP <AG>, shows a correlation coefficient close to unity (r=0.98); this high correlation indicates in both species, a highly similar <AG> distribution along both orthologs. The correlation coefficients for the remaining 11 tCP-profiles are, however, significantly lower. From data, we conclude that high percent NT similarity between orthologs do not imply higher number of shared tCPs (Figure 4c). similar study was performed for the remaining GPCRs from the dataset (see Table 2). Figure 4c also shows that the amount of shared tCPs in a notable fraction of orthologs varies linearly around the regression line with the percent tCP similarity; higher is the similarity, higher is the amount of shared tCPs and vice versa. For percent similarities higher than 70%, the convergence of data supports that two sequences 100% NT-identical would share 100% of tCPs, validating, consequently, the consistency of the process. Among others, many frizzled genes, playing a critical role in the regulation of cell polarity [48], meet the criteria set out above. These genes (Table 2), share a high number of tCPs (up to 12) and, therefore, high percent NT-identity with their mouse orthologs and therefore, a high degree of conservation interspecies. As the frizzled proteins are receptors for secreted Wnt proteins, the data support that the Wnt pathway would be evolutionarily ancient, as reported [49].

It can be observed that most GPCRs share less than 5 tCPs (Figure 4b), showing this region a large proportion of orthologs with high deviations around the mean and revealing remarkable composition patterns (Figure 4c); in this region, the number of shared tCPs and the percent NT identity do not correlate at all. In one case, the tCP compositional structures shown by orthologs in this region would not change notably at speciation events, being hard to find them because they would be often masked by the high percent of NT similarity observed as occur in ADRB2 (yellow point, Figure 4c). In the other case, the low NT percent similarity let us identify easily the underlying tCP-compositional structure shared by orthologs.

A case illustrating the precedent discussion would be of the human-mouse orthologs LGR4 and LGR5 both included in the group B of R-spondins [50]; they potentiate the Wnt/β-catenin signaling pathway and stem cell growth factors, being both, closely related to cancer [42, 51]. LGR4 and LGR5, despite their high structural and functional similarity [42, 52], have orthologs sharing very different number of tCPs (blue points, Figure 4c). The conserved compositional structure of LGR4, is subtler (2 conserved tCPs with its counterpart in mouse) than in LGR5 (8 conserved tCPs). None tCP conserved in LGR5 is shared by LGR4 (Figure 5a) despite the possibility to have a common ancestor in the early metazoan evolution [50] and suggesting, that LGR5 would be evolutionary ancient than LGR4 because many of changes occurring in LGR5, at speciation events, would be less than those occurring in LGR4. A functional feature common in LGR4 and LGR5 that differentiates from other GPCRs is the large N-terminal domain composed by 17 leucine-rich repeats, containing 24 amino acids each [42]. Centering the attention on this extracellular domain, we wonder on what and how many of these synonymous codons are in this large external region as well as in other protein domains. Extreme differences were observed between the synonymous codons encoding leucine in the tCP conserved and non-conserved regions of LGR4 and LGR5. Differences occur in all domains of the protein, being these ones clearly dissimilar between LGR4 and LGR5. It is the first time that this extreme synonymous codon preference for leucine is described in LGRs and could be explained by the degeneration of the tCP code (Table 1).

It is well known that changes in synonymous codons could have unexpected effects many of them linked with disease [53, 54]. Evidences reveal that synonymous codon selection in natural RNAs have evolved in response to selective pressures at both the RNA and protein levels [55]. In fact, high mRNA expression levels encode slow-evolving proteins in most organisms [56, 57]. It is known, that coding sequences accumulate non-synonymous substitutions at a slower rate than the synonymous substitution rate for correct protein folding and function [58]. In fact, Figure 3 indicates that compositional changes in GPCRs increase notably in tCP non-conserved regions relative to conserved regions, suggesting that optimal codons are significantly associated with conserved amino acids. This would be implying, as reported, that selection has placed optimal codons to diminish the consequences of translation errors in every organism and that optimal codon sets scored significantly higher than alternative sets in all organisms [59].

Lately, methods to improve protein production by using synonymous codon changes (codon-optimization) has been be reconsidered because it can affect the protein conformation and function [54, 59]. This paper provides an evolutionary approach to codon optimization by using the tCP-profile of ortholog sequences. The profile reveals the existence of evolutionary conserved tCPs in GPCR fragments, not previously detected. Results suggest that both, the set of optimal codons, and codon positions, interact to enhance translational precision warranting that codon bias can be significant for protein structure and function [58]. Data show, for each ortholog, that tCP conserved and non-conserved fragments show extreme synonymous codon preferences dissimilar to each other (Figure 6). Thus, the presence of the conserved compositional structure and the over usage of certain synonymous codons in those regions encourage us to suggest how we can modify the expression of some genes by synonymous codon optimization in order to avoid possible adverse effects on the protein conformation. Efforts are being done on the whole-genome synonymous codon substitution, a mechanism to create exclusive

organisms exhibiting genetic isolation and enhanced biological functions [60]. The repurposing of codons would be a strategy for enhancing genomes with functions not commonly found in nature [61, 62]. We think that the knowledge of these new conserved compositional structures would simplify notably the rational codon optimization design, circumventing or substituting synonymous codons in tCP conserved and non-conserved regions that would be important from the point of view of the structure and function.

Relative to ORs superfamily, it is categorized in at least nine divergent gene families, having little sequence similarity each and showing dramatic differences in size between species [63]. Moreover, during the vertebrate evolution the OR repertoire has expanded about 10-fold. Such diversification is likely to result from translocation, gene conversion, recombination, and/or duplication [64, 65]. Data herein reported suggest, however, that such divergence would be, to some extent, apparent and restricted to the reduced NT similarity between OR gene families. This suggests that although the NT similarity remains lower than that in non-OR-GPCRs, ORs retain the tCP compositional structures at speciation events when they sharing one or more tCPs. However, Figure 4b indicates that for $r \geq 0.85$, a significant fraction of GPCRs does not share any tCP between orthologs and that this proportion is notably higher for ORs than for non-OR-GPCRs, supporting data showing the low percent similarity observed between different OR families. The absence of resolved high-resolution structures in ORs have hampered an in-depth understanding into their molecular mechanism [11-13]. However, the tCP compositional structure suggests a new point of view to address the issue by supplying new conserved motifs previously unavailable and indicating that changes in these tCP motifs probably would affect the function of GPCRs. This occurs with many GPCRs of the rhodopsin family having several conserved transmembrane domains [66] whose study suggests that changes in those residues often result in misfolding of GPCRs [67]. In fact, some sites in the DNA are more prone to mutations than others and one would expect to find an excess of them that are polymorphic in not shared regions of orthologs. Receptor mutations in conserved transmembrane residues, however, selectively disrupt one pathway of a receptor coupled to multiple pathways [68-69]. For similar reasons we think that the tCP compositional structures preserved over millions of years [22 23] could be key evolutionary elements for fitness with potential to be used in mutation studies.

Now, we will discuss something about the amount and the type of orthologs detected when different algorithms are used in distinct databanks; For example, Figure 7 shows all mouse orthologs we have found of the human OR3A1 in both GeneCards and Ensembl. To test the response of the tCP method to this issue, we compared the tCP-profiles of all of them. The comparison was exclusively based in the tCP composition properties of the entire sequence length, not in local similarities. Results suggest that differences between databanks are consequence of the parameters used to discriminate orthologs, giving results, as in this case, very divergent between each other. The tCP method test results (Figure 7) are nearer to those obtained by Ensembl suggesting in this case that GeneCards would underestimate the number of mouse orthologs of OR3A1. In any case, analyzing carefully tCP-profiles of all mouse orthologs of OR3A1 proposed by the Ensembl algorithm, the tCP method suggests a dubious case (*Olfr1*) and a negative case (*Olfr10*). This occurs frequently and an analysis case-by-case would be required. Despite a significant fraction of ORs have lacking shared tCPs (Figure 4B), relevant evolutionary information could be obtained by tuning down the cut-off of tCP-profiles considering correlations lower but near the cut-off in the analysis. A scrutiny of such information could be useful to clarify data as those of the *Olfr1* and *Olfr10* (Figure 7) considered orthologs of OR3A1 by the Ensembl algorithm.

Highest NT percent similarities between orthologs not always reveal useful evolutionary information, masking, however, the shared tCP structures hidden behind the high NT similarity. The knowledge of tCPs conserved deduced from the analysis of tCP-profiles are useful to reveal such compositional structures. Thus, the mutation dynamics outside of the shared compositional structures (Figure 3) would explain why some GPCRs of the same functional family might evolve to different levels of complexity by adaptation to the environment because it is known that the divergence between orthologous would promote the evolution of novel functions. The data indicate that the shared tCP compositional structures between orthologs observed in evolutionary distant species would change very little being more reticent to mutate than those non shared ones [22, 23]. Table 2 supply information of many GPCRs to readers and for which we have only seen few examples, in order to illustrate the hypothesis proposed.

5. Conclusions

As a summary, we would highlight the following facts: i) identical NT sequences generate identical tCP sequences, but identical tCP sequences generate a plethora of different NT sequences due to the degeneration of the tCP code (Table 1); this open the possibility of having NT mismatches in DNA fragments tCP conserved letting to tCP conserved regions a certain degree of flexibility at NT level necessary to adaptation to the environment. However, NT mismatches in tCP conserved region are few in number compared with those observed in non-conserved tCP regions (Figure 3). ii) There are NT-misaligned fragments that can supplying useful evolutionary information, and the tCP-profile allows identifying them. iii) The degeneration of the tCP code (Table 1) would explain the extreme synonymous codon usage in the conserved compositional structures and in the non-conserved regions between ortholog-GPCRs (Figure 5). iv) The paper expands the concept of homology taking into account the percent similarity between tCP sequences containing, as a particular case, the NT percent similarity.

Appendix

Table 2. *List of human-mouse ortholog GPCRs analyzed. In parentheses, after the human OR, we indicated the name of the counterpart OR in mouse following the nomenclature of the international databases. Some human ORs have annotated more than one mouse ortholog and vice versa. The cut-off for conserved tCPs is r≥0.85.*

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
1	ADORA1	Adenosine receptor A1	4	72,8	4,9
2	ADORA2A	Adenosine receptor 2A	7	62	7,4
3	ADORA2B	Adenosine receptor 2B	3	71,4	4,2
4	ADORA3	Adenosine receptor 3	0	58,8	8,7
5	ADRA1A	Alpha-1A adrenergic receptor	7	70,4	2,9
6	ADRA1B	Alpha-1B adrenergic receptor	3	50,5	39,9
7	ADRA1D	Alpha-1D adrenergic receptor	5	61,6	11,7
8	ADRA2A	Alpha-2A adrenergic receptor	8	71,2	6,4
9	ADRA2B	Alpha-2B adrenergic receptor	5	63,6	8,7
10	ADRA2C	Alpha-2C adrenergic receptor	8	69,1	9,8
11	ADRB1	Beta-1 adrenergic receptor	5	75,2	8,1
12	ADRB2	Beta-2 adrenergic receptor	3	67,1	7,7
13	ADRB3	Beta-3 adrenergic receptor	5	59,1	12,7
14	AGTR1	Type-1 angiotensin II receptor	5	67,9	0,6
15	AGTR2	Type-2 angiotensin II receptor	6	76,6	2,7
16	AGTRL1	Apelin receptor A	4	71	3
17	AVPR1A	Vasopressin V1a receptor	5	62,2	8,9
18	AVPR1B	Vasopressin V1b receptor	4	63,5	8
19	AVPR2	Vasopressin V2 receptor	1	62,4	7
20	BAI1	Adhesion G protein-coupled receptor B1	6	71,3	5,3
21	BAI2	Adhesion G protein-coupled receptor B2	8	76,7	3,9
22	BAI3	Adhesion G protein-coupled receptor B3	8	80,9	1
23	BDKRB1	B1 bradykinin receptor	1	54,9	17,4
24	BDKRB2	B2 bradykinin receptor	1	60,6	11,8
25	BRS3	Bombesin receptor subtype-3	7	71,1	3,5
26	C3AR1	C3a anaphylatoxin chemotactic receptor	4	53,3	11,3
27	C5AR1	C5a anaphylatoxin chemotactic receptor 1	1	46,5	18,3
28	CALCR	Calcitonin receptor	2	51,9	18,6
29	CALCRL	Calcitonin gene-related peptide type 1 receptor	3	63,7	6,4
30	CASR	Extracellular calcium-sensing receptor	10	70,3	3,4
31	CCBP2	Atypical chemokine receptor 2	2	55,4	11,1
32	CCKAR	Cholecystokinin receptor type A	5	67,9	4,7
33	CCKBR	Gastrin/cholecystokinin receptor type B	5	65,7	7,4
34	CCR10	C-C chemokine receptor type 10	7	64,6	8,1
35	CCR2	C-C chemokine receptor type 2	3	52,3	16,5
36	CCR3	C-C chemokine receptor type 3	3	52,9	10,5
37	CCR4	C-C chemokine receptor type 4	3	61,6	6,4
38	CCR5	C-C chemokine receptor type 5	2	64,4	5,6
39	CCR6	C-C chemokine receptor type 6	1	50,7	12,7
40	CCR7	C-C chemokine receptor type 7	3	58,2	21,2
41	CCR8	C-C chemokine receptor type 8	3	53,5	14,9
42	CCR9	C-C chemokine receptor type 9	4	64,6	6,1
43	CCRL1	C-C chemokine receptor-like 1	2	62,3	1,9
44	CCRL2	C-C chemokine receptor-like 2	2	43,3	20,2
45	CCXCR1	Chemokine XC receptor 1	1	52,5	13,9
46	CD97	CD97 antigen	3	45,8	22,5
47	CELSR1	Cadherin EGF LAG seven-pass G-type receptor 1	7	59,1	9,8
48	CELSR2	Cadherin EGF LAG seven-pass G-type receptor 2	4	70,1	2,9
49	CELSR3	Cadherin EGF LAG seven-pass G-type receptor 3	7	69	3,9
50	CHRM1	Muscarinic acetylcholine receptor M1	7	78,3	1,5
51	CHRM2	Muscarinic acetylcholine receptor M2	8	74,7	1,7
52	CHRM3	Muscarinic acetylcholine receptor M3	7	69,3	4,8
53	CHRM4	Muscarinic acetylcholine receptor M4	7	72,8	2,4
54	CHRM5	Muscarinic acetylcholine receptor M5	8	73,7	1,1
55	CMKLR1	Chemokine-like receptor 1	4	60,9	11,3
56	CMKOR1	Atypical chemokine receptor 3	3	74	3,4
57	CNR2	GTP-binding nuclear protein GSP2/CNR2	2	57,1	9,5
58	CRHR1	Corticotropin-releasing factor receptor 1	8	77,9	1
59	CRHR2	Corticotropin-releasing factor receptor 2	2	39,5	45,9
60	CXCR2	C-X-C chemokine receptor type 2	0	56,2	9,5

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
61	CXCR3	C-X-C chemokine receptor type 3	5	61,1	8,3
62	CXCR4	C-X-C chemokine receptor type 4	3	69,8	2,7
63	CXCR5	C-X-C chemokine receptor type 5	2	60,1	7,8
64	CXCR6	C-X-C chemokine receptor type 6	0	57,9	6
65	CYSLT1	Cysteinyl leukotriene receptor 1	8	68,9	6,5
66	CYSLT2	Cysteinyl leukotriene receptor 2	2	48,4	21,6
67	DARC	Atypical chemokine receptor 1	4	45,1	16,7
68	DRD1	D (1) dopamine receptor	3	69,2	2
69	DRD2	D (2) dopamine receptor	6	71,2	7,7
70	DRD3	D (3) dopamine receptor	5	64	13,7
71	DRD4	D (4) dopamine receptor	4	60,8	18
72	DRD5	D (5) dopamine receptor	4	61,3	10,5
73	EBI2	G-protein coupled receptor 183	2	58,4	8,3
74	EDG1	Sphingosine 1-phosphate receptor 1	5	71,7	2,1
75	EDG2	Lysophosphatidic acid receptor 1	7	73,2	5,6
76	EDG3	Sphingosine 1-phosphate receptor 3	3	67,8	4,1
77	EDG4	Lysophosphatidic acid receptor 2	5	68,6	7,4
78	EDG5	Sphingosine 1-phosphate receptor 2	4	60	10,9
79	EDG6	Sphingosine 1-phosphate receptor 4	2	56,9	7,7
80	EDG7	Lysophosphatidic acid receptor 3	3	64,1	6
81	EDG8	Sphingosine 1-phosphate receptor 5	1	56,3	14,8
82	EDNRA	Endothelin-1 receptor	3	64,7	2
83	EDNRB	Endothelin receptor type B	6	68,5	4,2
84	ELTD1	Adhesion G protein-coupled receptor L4	4	54,9	14
85	EMR1	Adhesion G protein-coupled receptor E1	2	52,6	13,6
86	EMR4	Adhesion G protein-coupled receptor E4	0	37,1	39,9
87	F2RL1	Proteinase-activated receptor 2	3	59,2	9,6
88	F2RL2	Proteinase-activated receptor 3	2	52	7,8
89	F2RL3	Proteinase-activated receptor 4	2	52,5	15,7
90	FFA1R	Free fatty acid receptor 1	3	54,5	10
91	FFA2R	Free fatty acid receptor 2	4	51,8	18,8
92	FFA3R	Free fatty acid receptor 3	6	61,5	8,7
93	FPR1	fMet-Leu-Phe receptor	0	52,3	11,5
94	FPRL1	N-formyl peptide receptor 3	0	49,1	19,9
95	FSHR	Follicle-stimulating hormone receptor	11	68,8	2,8
96	FZD1	Frizzled-1	8	74,3	4,6
97	FZD2	Frizzled-2	8	78,7	5,7
98	FZD3	Frizzled-3	10	80,5	0,6
99	FZD4	Frizzled-4	12	80,8	1,5
100	FZD5	Frizzled-5	8	74,9	2,3
101	FZD6	Frizzled-6	4	49,3	29,2
102	FZD7	Frizzled-7	6	75,2	4,4
103	FZD8	Frizzled-8	7	71,8	7
104	FZD9	Frizzled-9	5	65,4	9,6
105	FZD10	Frizzled-10	7	68,3	3,2
106	GABABL	Probable G-protein coupled receptor 156	8	58,9	9,6
107	GABBR1	Gamma-aminobutyric acid type B receptor subunit 1	2	66	14,8
108	GABBR2	Gamma-aminobutyric acid type B receptor subunit 2	7	73,8	1,8
109	GALR1	Galanin receptor type 1	7	70,8	3,3
110	GALR2	Galanin receptor type 2	4	63,5	12
111	GALR3	Galanin receptor type 3	8	70,5	9,6
112	GCGR	Glucagon receptor	3	60,9	6,8
113	GHRHR	Growth hormone-releasing hormone receptor	0	63,4	5,1
114	GHSR	Growth hormone secretagogue receptor type 1	8	75.8	2.7
115	GIPR	Gastric inhibitory polypeptide receptor	4	60,2	12,5
116	GLP1R	Glucagon-like peptide 1 receptor	3	67	8
117	GLP2R	Glucagon-like peptide 2 receptor	1	61,3	11
118	GPBAR1	G-protein coupled bile acid receptor 1	3	62,9	3,3
119	GPR120	Free fatty acid receptor 4	7	63,9	10
120	GPR126	Adhesion G-protein coupled receptor G6	2	61,9	8,6
121	GPR18	G-protein coupled receptor 18	8	62,1	4,6
122	GPR19	G-protein coupled receptor 19	5	58,7	10,7
123	GPR20	G-protein coupled receptor 20	4	61,4	8,7
124	GPR21	G-protein coupled receptor 21	7	76,3	0
125	GPR22	G-protein coupled receptor 22	8	72,5	6,8
126	GPR23	G-protein coupled receptor 23	2	76	0,8
127	GPR25	G-protein coupled receptor 25	2	60,3	17,4

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
128	GPR26	G-protein coupled receptor 26	11	78,9	2,3
129	GPR27	G-protein coupled receptor 27	9	83,9	2,9
130	GPR30	G-protein coupled receptor 30	3	65,6	10,2
131	GPR31	G-protein coupled receptor 31	2	48,9	13,2
132	GPR32	G-protein coupled receptor 32	0	43,1	19,9
133	GPR33	G-protein coupled receptor 33	4	56,5	11
134	GPR34	G-protein coupled receptor 34	3	68,3	5,9
135	GPR35	G-protein coupled receptor 35	1	55	18,3
136	GPR37	G-protein coupled receptor 37	7	60,4	10,2
137	GPR37L1	Prosaposin receptor GPR37L1	2	52,9	29,4
138	GPR39	G-protein coupled receptor 39	4	50	10,7
139	GPR4	G-protein coupled receptor 4	4	69,5	6,1
140	GPR44	G-protein coupled receptor 44	2	60,3	10,7
141	GPR45	G-protein coupled receptor 45	6	66,2	5,8
142	GPR50	G-protein coupled receptor 50	5	59,2	12,6
143	GPR52	G-protein coupled receptor 52	4	66,2	15
144	GPR55	G-protein coupled receptor 55	1	60,1	7,5
145	GPR56	G-protein coupled receptor 56	6	60,2	10,1
146	GPR6	G-protein coupled receptor 6	5	68,4	6,1
147	GPR61	G-protein coupled receptor 61	8	77,8	3,1
148	GPR62	G-protein coupled receptor 62	3	62,6	12,6
149	GPR63	G-protein coupled receptor 63	5	69,1	3,8
150	GPR64	G-protein coupled receptor 64	5	63,2	11,8
151	GPR65	G-protein coupled receptor 65	2	54,7	12,7
152	GPR68	G-protein coupled receptor 68	4	68,1	7,1
153	GPR75	G-protein coupled receptor 75	5	68,1	3,8
154	GPR77	G-protein coupled receptor 77	1	43,6	25,8
155	GPR79	G-protein coupled receptor 79	2	48,5	21,1
156	GPR81	Hydroxycarboxylic acid receptor 1	3	59,3	10
157	GPR82	Probable G-protein coupled receptor 82	4	68,3	3,1
158	GPR83	Probable G-protein coupled receptor 83	1	64,8	11
159	GPR84	G-protein coupled receptor 84	3	64,3	3,9
160	GPR85	Probable G-protein coupled receptor 85	9	87,8	0,3
161	GPR87	G-protein coupled receptor 87	3	64,4	4
162	GPR88	Probable G-protein coupled receptor 88	8	77,3	3,7
163	GPR92	Lysophosphatidic acid receptor 5	6	58,4	15,2
164	GPR97	Adhesion G protein-coupled receptor G3	3	53,4	12,9
165	GPR101	Probable G-protein coupled receptor 101	5	47	25,8
166	GPR103	Pyroglutamylated RFamide peptide receptor	4	52,8	24,8
168	GPR107	Protein GPR107	5	63	7,8
169	GPR108	G-protein coupled receptor 108	3	58,8	13,1
170	GPR109A	Hydroxycarboxylic acid receptor 2	1	64,6	7,7
171	GPR110	Adhesion G-protein coupled receptor F1	1	50,7	15,7
172	GPR111	Adhesion G-protein coupled receptor F2	1	43,8	31,2
173	GPR112	Adhesion G-protein coupled receptor G4	4	47,3	19
174	GPR113	Adhesion G-protein coupled receptor F3	2	42,6	26,8
175	GPR114	Adhesion G-protein coupled receptor G5	2	46,7	21
176	GPR115	Adhesion G protein-coupled receptor F4	4	49,7	15,8
177	GPR116	Adhesion G protein-coupled receptor F5	2	50,3	18,6
178	GPR119	Glucose-dependent insulinotropic receptor	3	66,6	2,9
179	GPR120	Free fatty acid receptor 4	6	63,9	10
180	GPR123	Adhesion G protein-coupled receptor A1	0	25,1	62,7
181	GPR124	Adhesion G protein-coupled receptor A2	5	64,2	8,9
182	GPR125	Adhesion G protein-coupled receptor A3	0	60,4	12,5
183	GPR126	Adhesion G-protein coupled receptor G6	2	61,9	8,6
184	GPR128	Adhesion G-protein coupled receptor G7	2	53,2	13,1
185	GPR132	Probable G-protein coupled receptor 132	2	50,7	20,7
186	GPR133	Adhesion G-protein coupled receptor D1	0	50,5	22,5
187	GPR135	Probable G-protein coupled receptor 135	5	59,6	15,2
188	GPR137B	Integral membrane protein GPR137B	10	72,6	7
189	GPR137	Integral membrane protein GPR137	4	58,7	24,4
190	GPR137C	Integral membrane protein GPR137C	5	42,9	44,7
191	GPR139	Probable G-protein coupled receptor 139	5	74,8	3,5
192	GPR141	Probable G-protein coupled receptor 141	0	60,4	5,4
193	GPR142	Probable G-protein coupled receptor 142	2	46	29,2
194	GPR143	G-protein coupled receptor 143	1	44,7	31,9
195	GPR144	Adhesion G-protein coupled receptor D2	1	32	46,9

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
196	GPR146	Probable G-protein coupled receptor 146	3	54,2	14,7
197	GPR149	Probable G-protein coupled receptor 149	6	36,5	45,9
198	GPR150	Probable G-protein coupled receptor 150	3	53,2	18,2
199	GPR151	Probable G-protein coupled receptor 151	5	62,4	6
200	GPR152	Probable G-protein coupled receptor 152	6	53,1	19,6
201	GPR153	Probable G-protein coupled receptor 153	7	63,6	8,8
202	GPR154	Neuropeptide S receptor	3	60,4	16
203	GPR157	G-protein coupled receptor 157	4	63,3	7,7
204	GPR158L1	Probable G-protein coupled receptor 179	9	54,7	12,6
205	GPR160	Probable G-protein coupled receptor 160	0	41,4	19,5
206	GPR161	G-protein coupled receptor 161	1	34,6	53,1
207	GPR162	Probable G-protein coupled receptor 162	7	74,1	2,1
208	GPR165	G protein-coupled receptor 165	0	51,4	18,8
209	GPR166P	G protein-coupled receptor 166p	0	45,8	21
210	GPR171	Probable G-protein coupled receptor 171	4	65,9	3,4
211	GPR172B	Solute carrier family 52, riboflavin transporter, member 1	2	62,8	9,4
212	GPR173	Probable G-protein coupled receptor 173	10	86,2	1,8
213	GPR174	Probable G-protein coupled receptor 174	2	69,2	3,6
214	GPR175	Transmembrane protein adipocyte-associted 1	0	58,5	20,4
215	GPR176	G-protein coupled receptor 176	8	71,3	5
216	GPR177	Protein wntless homolog	6	74,6	1,1
217	GPR178	Transmembrane protein 181	2	44,7	32,7
218	GPRC5A	Retinoic acid-induced protein 3	3	57,6	8,4
219	GPRC5B	G-protein coupled receptor family C group 5 member B	6	67,9	6,6
220	GPRC5C	G-protein coupled receptor family C group 5 member C	8	66	5,9
221	GPRC5D	G-protein coupled receptor family C group 5 member D	2	58,9	5,9
222	GPRC6A	G-protein coupled receptor family C group 6 member A	1	62,9	4,3
223	GRM1	Metabotropic glutamate receptor 1	9	71,2	3,5
224	GRM2	Metabotropic glutamate receptor 2	3	73,3	2
225	GRM3	Metabotropic glutamate receptor 3	9	74,6	1,1
226	GRM4	Metabotropic glutamate receptor 4	4	74,4	3,2
227	GRM5	Metabotropic glutamate receptor 5	4	59,7	23,7
228	GRM6	Metabotropic glutamate receptor 6	0	69,9	4,6
229	GRM7	Metabotropic glutamate receptor 7	10	79	1,4
230	GRM8	Metabotropic glutamate receptor 8	6	78,3	1,8
231	GRPR	Gastrin-releasing peptide receptor	5	71,1	4,8
232	HCRTR1	Orexin receptor type 1	6	71,2	7
233	HCRTR2	Orexin receptor type 2	6	67,8	7
234	HRH1	Histamine H1 receptor	6	60,1	9,6
235	HRH2	Histamine H2 receptor	4	43,9	38,5
236	HRH3	Histamine H3 receptor	5	72,9	5,7
237	HRH4	Histamine H4 receptor	3	51,4	12
238	HTR1A	5-hydroxytryptamine receptor 1A	7	66,7	6,3
239	HTR1B	5-hydroxytryptamine receptor 1B	7	75,8	2,5
240	HTR1D	5-hydroxytryptamine receptor 1D	5	66,8	4,6
241	HTR1F	5-hydroxytryptamine receptor 1F	5	71,5	0,5
242	HTR2B	5-hydroxytryptamine receptor 2B	5	61,8	6,4
243	HTR5A	5-hydroxytryptamine receptor 5A	4	62,6	2,9
244	OPRK1	Kappa-type opioid receptor	7	67,6	2,9
245	OPRL1	Nociceptin receptor	5	90,3	9,7
246	PRRP	Prolactin-releasing peptide receptor	3	28,7	45
247	PTGFR	Prostaglandin F2-alpha receptor	3	49,7	21,9
248	TAAR3	Putative trace amine-associated receptor 3	4	65,6	1,2
249	TACR2	Substance-K receptor	2	61,8	11
250	TM7SF3	Transmembrane 7 superfamily member 3	3	58,3	3,6
251	TM7SF4	Dendritic cell-specific transmembrane Protein	3	57,7	7,2
252	HTR2A	5-hydroxytryptamine receptor 2A	7	72	1,2
253	HTR2B	5-hydroxytryptamine receptor 2B	5	61,8	6,4
254	HTR2C	5-hydroxytryptamine receptor 2C	5	70	4,8
255	HTR4	5-hydroxytryptamine receptor 4	8	78,5	1,4
256	HTR5A	5-hydroxytryptamine receptor 5A	4	62,6	2,9
257	HTR5B	5-hydroxytryptamine receptor 5B	1	55,7	19
258	HTR6	5-hydroxytryptamine receptor 6	3	69,5	5,6
259	HTR7	5-hydroxytryptamine receptor 7	11	74,5	4,2
260	KISS1R	KiSS-1 receptor	3	51	26,1
261	LGR4	Leucine rich repeat containing G-protein coupled receptor 4	2	73,1	3,6
262	LGR5	Leucine rich repeat containing G-protein coupled receptor 5	8	64,2	4,6

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
263	LGR6	Leucine rich repeat containing G-protein coupled receptor 6	8	64,1	10,4
264	LHCGR	Lutropin choriogonadotropic hormone receptor	4	63,3	7
265	LPHN1	Adhesion G protein-coupled receptor L1	4	56	20,3
266	LPHN2	Adhesion G protein-coupled receptor L2	9	71,7	4,8
267	LPHN3	Adhesion G protein-coupled receptor L3	9	72,1	6,5
268	LTB4R	Leukotriene B4 receptor 1	5	57,6	14,9
269	LTB4R2	Leukotriene B4 receptor 2	3	61,7	15
270	TMEM185A	Transmembrane protein 185A	7	71,5	9,2
271	TMEM185B	Transmembrane protein 185B	9	73	0,8
272	PAQR5	Membrane progestin receptor gamma	4	72,8	2,6
273	PAQR7	Membrane progestin receptor alpha	1	62,9	5,5
274	PAQR8	Membrane progestin receptor beta	3	68,2	1,7
275	UTS2R	Urotensin-2 receptor	2	52,9	14,4
276	TAAR1	Trace amine-associated receptor 1	3	55,3	7,7
277	TAAR2	Trace amine-associated receptor 2	3	58,4	11,3
278	TAAR3	Trace amine-associated receptor 3	4	65,6	1,2
279	TAAR5	Trace amine-associated receptor 5	6	67,1	2,7
280	TAAR6	Trace amine-associated receptor 6	7	64,8	3
281	TAAR9	Trace amine-associated receptor 9	5	62,9	3,6
282	ADMR	G-protein coupled receptor 182	3	59,7	11,8
283	MAS1	MAS1 oncogene	3	68	2,9
284	MC1R	Melanocyte-stimulating hormone receptor	2	59	6,3
285	MC2R	Adrenocorticotropic hormone receptor	3	63,4	6,2
286	MC3R	Melanocortin receptor 3	6	72	3,2
287	MC4R	Melanocortin receptor 4	7	72,4	0,6
288	MC5R	Melanocortin receptor 5	4	61,1	4
289	MRGPRD	Mas-related G-protein coupled receptor D	1	49,1	17,8
290	MRGPRE	Mas-related G-protein coupled receptor E	2	57,9	14,5
291	MRGPRF	Mas-related G-protein coupled receptor F	4	63,3	8,6
292	MRGPRG	Mas-related G-protein coupled receptor G	0	42,5	22,3
293	MTNR1A	Melatonin receptor type 1A	2	40,4	33,2
294	MTNR1B	Melatonin receptor type 1B	3	56,1	13,7
295	NMUR1	Neuromedin-U receptor 1	1	48,1	23,7
296	NMUR2	Neuromedin-U receptor 2	1	59,7	7,2
297	NPFF1	Neuropeptide FF receptor 1	7	66,7	10,7
298	NPFF2	Neuropeptide FF receptor 2	6	49,9	21,5
299	NPY1R	Neuropeptide Y receptor type 1	1	28,3	37,4
300	NPY2R	Neuropeptide Y receptor type 2	3	71,3	2,9
301	NPY5R	Neuropeptide Y receptor type 5	4	63,2	7,1
302	NTSR1	Neurotensin receptor type 1	3	63,3	9,4
303	NTSR2	Neurotensin receptor type 2	5	63,5	11,4
304	OPN3	Opsin-3	9	66,5	4,8
305	OPN4	Opsin-4	3	57,3	17,2
306	OPRD1	Delta-type opioid receptor	7	74,2	5,6
307	OPRK1	Kappa-type opioid receptor	7	67,6	2,9
308	OPRL1	Nociceptin receptor	10	90,3	9,7
309	OPRM1	Mu-type opioid receptor	2	53,9	23,7
310	OR2B11	Olfactory receptor 2B11 (Olfr222)	5	64	3,2
311	OR2C3	Olfactory receptor 2C3 (Olfr1389)	1	37	33,3
312	OR3A1	Olfactory receptor 3A1 (Olfr10)	0	29,9	38,2
313	OR3A1	Olfactory receptor 3A1 (Olfr139)	1	50,3	11,9
314	OR3A1	Olfactory receptor 3A1 (Olfr399)	2	52,9	12,4
315	OR3A1	Olfactory receptor 3A1 (Olfr402)	4	63,9	3,7
316	OR3A1	Olfactory receptor 3A1 (Olfr410)	5	66,2	1,3
317	OR3A1	Olfactory receptor 3A1 (Olfr411)	2	63,8	3,9
318	OR3A1	Olfactory receptor 3A1 (Olfr1115)	1	32,7	31,1
319	OR3A1	Olfactory receptor 3A1 (Olfr1)	2	33,5	39,9
320	OR2T8	Olfactory receptor 2T8 (Olfr314)	2	54	12,1
321	OR2T12	Olfactory receptor 2T12 (Olfr314)	2	54,6	8,8
322	OR2A14	Olfactory receptor 2A14 (Olfr444)	6	62,8	5,6
323	OR2C1	Olfactory receptor 2C1 (Olfr15)	3	60,1	6,2
324	OR13J1	Olfactory receptor 13J1 (Olfr71)	1	67,1	2,9
325	OR13A1	Olfactory receptor 13A1 (Olfr211)	3	63	10,4
326	OR10S1	Olfactory receptor 10S1 (Olfr982)	3	63	6,5
327	OR5C1	Olfactory receptor 5C1 (Olfr368)	3	63,1	7,2
328	OR6F1	Olfactory receptor 6F1 (Olfr308)	1	58,7	1,9
329	OR2Y1	Olfactory receptor 2Y1 (Olfr1388)	4	62,5	6,4

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
330	OR11L1	Olfactory receptor 11L1 (Olfr323)	4	55,3	16,9
331	OR4D2	Olfactory receptor 4D1 (Olfr462)	5	67,9	3,8
332	OR5A1	Olfactory receptor 5A1 (Olfr76)	0	63,7	1,7
333	OR2H2	Olfactory receptor 2H2 (Olfr93)	6	63,6	7
334	OR10Q1	Olfactory receptor 10Q1 (Olfr495)	0	27,9	44,5
335	OR56A4	Olfactory receptor 5A4 (Olfr684)	2	57,9	18
336	OR10G3	Olfactory receptor 10G3 (Olfr1512)	0	63,7	0,4
337	OR10G2	Olfactory receptor 10G2 (Olfr1510)	3	69,6	3,7
338	OR2A5	Olfactory receptor 2A5 (Olfr448)	5	66,6	3,9
339	OR2AJ1	Olfactory receptor 2AJ1 (Olfr170)	3	55,6	9,4
340	OR2L8	Olfactory receptor 2L8 (Olfr167)	3	62,1	0,9
341	OR2L5	Olfactory receptor 2L5 (Olfr167)	5	60,9	2,3
342	OR2L2	Olfactory receptor 2L2 (Olfr167)	4	58,3	3,6
343	OR2B6	Olfactory receptor 2B6 (Olfr11)	4	58,8	2,1
344	OR2W1	Olfactory receptor 2W1 (Olfr263)	6	67,2	4,2
345	OR2J3	Olfactory receptor 2J3 (Olfr137)	2	57,1	3,9
346	OR2F2	Olfactory receptor 2F2 (Olfr453)	5	59,8	4,3
347	OR2F2	Olfactory receptor 2F2 (Olfr452)	7	64,7	3,9
348	OR2F2	Olfactory receptor 2F2 (Olfr438)	5	64,2	3,7
349	OR2A25	Olfactory receptor 2A25 (Olfr447)	5	60,8	4,2
350	OR2K2	Olfactory receptor 2K2 (Olfr267)	6	69,2	1,2
351	OR2D2	Olfactory receptor 2D2 (Olfr715)	2	61,5	6,1
352	OR2D3	Olfactory receptor 2D3 (Olfr716)	3	59,6	7,9
353	OR56B1	Olfactory receptor 56B1 (Olfr657)	3	63	5,9
354	OR56B1	Olfactory receptor 56B1 (Olfr504)	4	61,9	6,6
355	OR52E2	Olfactory receptor 52E2 (Olfr589)	2	61,3	3,5
356	OR52A5	Olfactory receptor 52A5 (Olfr68)	2	59,7	3,8
357	OR52A5	Olfactory receptor 52A5 (Olfr69)	1	58,5	5,1
358	OR52N4	Olfactory receptor 52N4 (Olfr658)	3	68	3,5
359	OR52N4	Olfactory receptor 52N4 (Olfr503)	2	64,4	3,7
360	OR52N5	Olfactory receptor 52N5 (Olfr669)	3	65,4	3,8
361	OR52N1	Olfactory receptor 52N1 (Olfr664)	1	62,2	7,6
362	OR52E6	Olfactory receptor 52E6 (Olfr671)	2	63,8	5,2
363	OR52E6	Olfactory receptor 52E6 (Olfr675)	3	63,1	8,3
364	OR52E8	Olfactory receptor 52E8 (Olfr671)	2	64,9	4,4
365	OR52E8	Olfactory receptor 52E8 (Olfr675)	2	63,1	5,1
366	OR52E4	Olfactory receptor 52E4 (Olfr677)	1	65,1	2,3
367	OR51F2	Olfactory receptor 51F2 (Olfr568)	1	62	10
368	OR51A7	Olfactory receptor 51A7 (Olfr576)	5	68,7	2,7
369	OR51A4	Olfactory receptor 51A4 (Olfr586)	1	44,8	14,3
370	OR51A2	Olfactory receptor 51A2 (Olfr586)	1	44,9	14,3
371	OR51B4	Olfactory receptor 51B4 (Olfr66)	1	56,4	9,1
372	OR51B6	Olfactory receptor 51B6 (Olfr65)	2	66,4	1,2
373	OR51M1	Olfactory receptor 51M1 (Olfr631)	1	61,3	6,1
374	OR13G1	Olfactory receptor 13G1 (Olfr309)	3	63,5	4,3
375	OR13F1	Olfactory receptor 13F1 (Olfr275)	4	62	2,3
376	OR13C3	Olfactory receptor 13C3 (Olfr273)	3	59,6	13,8
377	OR13C8	Olfactory receptor 13C8 (Olfr272)	1	60,9	6,4
378	OR13D1	Olfactory receptor 13D1 (Olfr270)	3	60,1	11,8
379	OR13H1	Olfactory receptor 13H1 (Olfr1235)	1	33	28,2
380	OR12D3	Olfactory receptor 12D3 (Olfr109)	1	56,3	6,5
381	OR11H6	Olfactory receptor 11H6 (Olfr745)	4	65,1	7,4
382	OR11H4	Olfactory receptor 11H4 (Olfr747)	2	66,1	7
383	OR11H4	Olfactory receptor 11H4 (Olfr749)	4	66,8	6,6
384	OR10X1	Olfactory receptor 10X1 (Olfr248)	1	61,5	1,1
385	OR10X1	Olfactory receptor 10X1 (Olfr417)	3	65,6	1,5
386	OR10J5	Olfactory receptor 10J5 (Olfr16)	4	66,1	3,6
387	OR10A3	Olfactory receptor 10A3 (Olfr514)	1	55	8,1
388	OR10A3	Olfactory receptor 10A3 (Olfr517)	0	56,4	7,1
389	OR10A3	Olfactory receptor 10A3 (Olfr519)	3	58,3	3,8
390	OR10A3	Olfactory receptor 10A3 (Olfr516)	4	61,4	3,3
391	OR10A3	Olfactory receptor 10A3 (Olfr512)	2	56,5	7,8
392	OR10A3	Olfactory receptor 10A3 (Olfr518)	2	57,5	10,8
393	OR10V1	Olfactory receptor 10V1 (Olfr420)	7	70,5	0,8
394	OR9A4	Olfactory receptor 9A4 (Olfr460)	5	66,2	2,4
395	OR9A2	Olfactory receptor 9A2 (Olfr459)	7	63,2	8,5
396	OR9G4	Olfactory receptor 9G4 (Olfr1006)	5	65,1	5,3

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
397	OR9G4	Olfactory receptor 9G4 (Olfr1015)	4	58,1	9,3
398	OR9K2	Olfactory receptor 9K2 (Olfr825)	4	61,3	10,2
399	OR9K2	Olfactory receptor 9K2 (Olfr926)	0	32,5	31,7
400	OR8I2	Olfactory receptor 8I2 (Olfr1104)	4	62,9	1,2
401	OR8H3	Olfactory receptor 8H3 (Olfr1100)	3	53,7	12,3
402	OR8H3	Olfactory receptor 8H3 (Olfr1099)	1	52,1	10,9
403	OR8H3	Olfactory receptor 8H3 (Olfr1098)	2	52,1	9,4
404	OR8H3	Olfactory receptor 8H3 (Olfr1097)	2	53,4	7,3
405	OR8J3	Olfactory receptor 8J3 (Olfr1045)	3	52,8	13,8
406	OR8H1	Olfactory receptor 8H1 (Olfr1100)	1	54,9	9,6
407	OR8H1	Olfactory receptor 8H1 (Olfr1099)	2	53,4	11
408	OR8H1	Olfactory receptor 8H1 (Olfr1098)	2	53,7	9,6
409	OR8H1	Olfactory receptor 8H1 (Olfr1097)	2	62,3	3,2
410	OR8J1	Olfactory receptor 8J1 (Olfr1045)	4	59,5	2,9
411	OR8U1	Olfactory receptor 8U1 (Olfr52)	2	63,3	4,7
412	OR8D4	Olfactory receptor 8D4 (Olfr895)	1	38,3	20,9
413	OR8G5	Olfactory receptor 8G5 (Olfr146)	1	49,6	19,9
414	OR8G5	Olfactory receptor 8G5 (Olfr935)	0	54,3	15
415	OR8G5	Olfactory receptor 8G5 (Olfr936)	0	43,3	27,6
416	OR8D2	Olfactory receptor 8D2 (Olfr924)	2	64,1	1,6
417	OR8D2	Olfactory receptor 8D2 (Olfr926)	2	61,4	2
418	OR8B2	Olfactory receptor 8B2 (Olfr918)	2	60,5	7,5
419	OR8B2	Olfactory receptor 8B2 (Olfr147)	2	63,8	4,3
420	OR8B3	Olfactory receptor 8B3 (Olfr918)	2	60,5	7,5
421	OR8B3	Olfactory receptor 8B3 (Olfr147)	3	72,7	7,3
422	OR8B4	Olfactory receptor 8B4 (Olfr878)	1	63,9	4,1
423	OR7A10	Olfactory receptor 7A10 (Olfr1353)	0	51,9	3,2
424	OR7A10	Olfactory receptor 7A10 (Olfr1352)	2	52,6	8,1
425	OR7A10	Olfactory receptor 7A10 (Olfr19)	0	54,5	6,3
426	OR7A10	Olfactory receptor 7A10 (Olfr57)	3	55,7	12,9
427	OR7A10	Olfactory receptor 7A10 (Olfr1355)	0	47,1	21,4
428	OR7A10	Olfactory receptor 7A10 (Olfr1351)	2	53,5	16,2
429	OR7A10	Olfactory receptor 7A10 (Olfr8)	0	46,7	21,6
430	OR7A10	Olfactory receptor 7A10 (Olfr1354)	0	47,2	18,8
431	OR6K2	Olfactory receptor 6K2 (Olfr420)	4	65,4	4,1
432	OR6K3	Olfactory receptor 6K3 (Olfr424)	0	53,1	11
433	OR6K6	Olfactory receptor 6K6 (Olfr231)	2	57,3	13
434	OR6N2	Olfactory receptor 6N2 (Olfr231)	0	35,7	24,4
435	OR6B1	Olfactory receptor 6B1 (Olfr449)	4	68,1	0,6
436	OR6C75	Olfactory receptor 6C75 (Olfr790)	5	69,3	3,3
437	OR6C2	Olfactory receptor 6C2 (Olfr791)	3	69,9	0
438	OR5K1	Olfactory receptor 5K1 (Olfr173)	5	63,1	6,5
439	OR5K2	Olfactory receptor 5K2 (Olfr173)	5	62,8	5,4
440	OR5K1	Olfactory receptor 5K1 (Olfr172)	3	63,8	1,6
441	OR5K2	Olfactory receptor 5K2 (Olfr172)	5	61,4	3,7
442	OR5P2	Olfactory receptor 5P2 (Olfr502)	2	57,8	12,2
443	OR5P2	Olfactory receptor 5P2 (Olfr506)	4	55,2	13,7
444	OR5P2	Olfactory receptor 5P2 (Olfr482)	2	57,2	7,8
445	OR5P2	Olfactory receptor 5P2 (Olfr507)	3	57,2	12,2
446	OR5P2	Olfactory receptor 5P2 (Olfr483)	4	54,8	14,1
447	OR5P2	Olfactory receptor 5P2 (Olfr493)	1	57,4	7,8
448	OR5P2	Olfactory receptor 5P2 (Olfr478)	5	59,2	9,4
449	OR5P2	Olfactory receptor 5P2 (Olfr497)	4	54	12,9
450	OR5P2	Olfactory receptor 5P2 (Olfr469)	1	56,2	10,7
451	OR5P2	Olfactory receptor 5P2 (Olfr487)	2	50,2	15,6
452	OR5P2	Olfactory receptor 5P2 (Olfr486)	2	49,5	15,6
453	OR5P2	Olfactory receptor 5P2 (Olfr510)	4	55,1	12,2
454	OR5P2	Olfactory receptor 5P2 (Olfr488)	4	51,3	12,2
455	OR5P2	Olfactory receptor 5P2 (Olfr485)	3	52,7	14,6
456	OR5P2	Olfactory receptor 5P2 (Olfr492)	2	53,8	11,6
457	OR5P2	Olfactory receptor 5P2 (Olfr470)	2	54,9	12
458	OR5P2	Olfactory receptor 5P2 (Olfr494)	4	54,4	11,3
459	OR5P2	Olfactory receptor 5P2 (Olfr490)	1	52,8	10,1
460	OR5P2	Olfactory receptor 5P2 (Olfr484)	4	57,2	8,8
461	OR5P2	Olfactory receptor 5P2 (Olfr495)	4	54,2	8,8
462	OR5P3	Olfactory receptor 5P3 (Olfr480)	1	53,5	14,5
463	OR5P3	Olfactory receptor 5P3 (Olfr508)	2	59,6	7,9

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
464	OR5L1	Olfactory receptor 5L1 (Olfr1156)	1	52	15,9
465	OR5L1	Olfactory receptor 5L1 (Olfr1157)	1	60,9	9,9
466	OR5L2	Olfactory receptor 5L2 (Olfr1156)	1	52	15,9
467	OR5L2	Olfactory receptor 5L2 (Olfr1157)	1	60,9	9,9
468	OR5D16	Olfactory receptor 5D16 (Olfr1155)	2	59,9	7,8
469	OR5W2	Olfactory receptor 5W2 (Olfr1152)	2	52,7	9
470	OR5W2	Olfactory receptor 5W2 (Olfr1153)	2	51,7	7,6
471	OR5W2	Olfactory receptor 5W2 (Olfr1140)	2	52,1	12,4
472	OR5W2	Olfactory receptor 5W2 (Olfr1141)	3	51,7	8,5
473	OR5W2	Olfactory receptor 5W2 (Olfr1138)	3	56,7	5,5
474	OR5W2	Olfactory receptor 5W2 (Olfr1137)	0	48,4	9,8
475	OR5W2	Olfactory receptor 5W2 (Olfr1136)	1	47,6	12,3
476	OR5W2	Olfactory receptor 5W2 (Olfr1135)	1	54,1	8,2
477	OR5I1	Olfactory receptor 5I1 (Olfr152)	3	59,2	6,5
478	OR5J2	Olfactory receptor 5J2 (Olfr1052)	2	55,3	7,4
479	OR5T2	Olfactory receptor 5T2 (Olfr1101)	3	49,7	19,3
480	OR5T2	Olfactory receptor 5T2 (Olfr1095)	2	48,9	16,9
481	OR5T2	Olfactory receptor 5T2 (Olfr1086)	3	50,4	17,7
482	OR5T2	Olfactory receptor 5T2 (Olfr141)	4	50,2	12,8
483	OR5T3	Olfactory receptor 5T3 (Olfr1093)	4	56,6	9
484	OR5T1	Olfactory receptor 5T1 (Olfr1093)	3	56,3	3,8
485	OR5M9	Olfactory receptor 5M9 (Olfr1036)	3	62,7	3,8
486	OR5M9	Olfactory receptor 5M9 (Olfr1034)	3	63,6	2,2
487	OR5M3	Olfactory receptor 5M3 (Olfr1032)	4	66,4	1,8
488	OR5M3	Olfactory receptor 5M3 (Olfr1033)	3	66,1	9,5
489	OR5M8	Olfactory receptor 5M8 (Olfr1031)	1	58,6	10
490	OR5M11	Olfactory receptor 5M11 (Olfr1028)	3	62,9	7,6
491	OR5M11	Olfactory receptor 5M11 (Olfr1029)	2	63,4	5,2
492	OR5M10	Olfactory receptor 5M10 (Olfr1023)	4	62,6	5,2
493	OR5M10	Olfactory receptor 5M10 (Olfr1022)	6	62,7	1,7
494	OR5AP2	Olfactory receptor 5AP2 (Olfr1020)	7	62,6	5
495	OR5AR1	Olfactory receptor 5AP2 (Olfr1019)	5	71,9	0,6
496	OR5AK2	Olfactory receptor 5AK2 (Olfr994)	0	55,5	9,3
497	OR5AK2	Olfactory receptor 5AK2 (Olfr995)	1	55,2	5,8
498	OR5B3	Olfactory receptor 5B3 (Olfr1461)	0	36,5	25,9
499	OR5B3	Olfactory receptor 5B3 (Olfr1458)	1	56,9	6,1
500	OR5B3	Olfactory receptor 5B3 (Olfr1467)	1	55,7	8,8
501	OR5B3	Olfactory receptor 5B3 (Olfr1469)	2	59,7	6,5
502	OR5B3	Olfactory receptor 5B3 (Olfr1459)	2	57,4	7,3
503	OR5B3	Olfactory receptor 5B3 (Olfr1462)	0	52,6	8,9
504	OR5B3	Olfactory receptor 5B3 (Olfr1446)	0	51,1	7,4
505	OR5B3	Olfactory receptor 5B3 (Olfr1453)	0	52,9	6,6
506	OR5B3	Olfactory receptor 5B3 (Olfr1447)	0	49,8	12
507	OR5B3	Olfactory receptor 5B3 (Olfr1454)	1	54,1	8,9
508	OR5B3	Olfactory receptor 5B3 (Olfr1457)	0	54,7	7,2
509	OR5B3	Olfactory receptor 5B3 (Olfr1463)	2	52,6	7,6
510	OR5B2	Olfactory receptor 5B2 (Olfr1465)	1	55,7	5,1
511	OR5B2	Olfactory receptor 5B2 (Olfr1451)	1	57,2	6,6
512	OR5B12	Olfactory receptor 5B12 (Olfr1445)	5	62,7	1,5
513	OR5B12	Olfactory receptor 5B12 (Olfr1448)	5	62,2	3,1
514	OR5AN1	Olfactory receptor 5AN1 (Olfr235)	0	60,3	9,9
515	OR5AN1	Olfactory receptor 5AN1 (Olfr1436)	0	56,1	7,4
516	OR5AN1	Olfactory receptor 5AN1 (Olfr262)	2	58,8	7,1
517	OR5AN1	Olfactory receptor 5AN1 (Olfr1431)	6	58,8	9,7
518	OR5AN1	Olfactory receptor 5AN1 (Olfr1434)	1	61,2	6,9
519	OR5AN1	Olfactory receptor 5AN1 (Olfr1437)	1	58,7	7,5
520	OR4F5	Olfactory receptor 4F5 (Olfr1289)	3	58,5	6,3
521	OR4F29	Olfactory receptor 4F29 (Olfr1303)	1	61,9	5,8
522	OR4F21	Olfactory receptor 4F21 (Olfr1303)	2	61,8	5,8
523	OR4B1	Olfactory receptor 4B1 (Olfr1272)	1	66,6	2,7
524	OR4B1	Olfactory receptor 4B1 (Olfr1271)	1	60,2	5,5
525	OR4B1	Olfactory receptor 4B1 (Olfr142)	2	63,2	5,1
526	OR4B1	Olfactory receptor 4B1 (Olfr1270)	5	65,7	5,6
527	OR4B1	Olfactory receptor 4B1 (Olfr32)	3	63,9	2,2
528	OR4C3	Olfactory receptor 4C3 (Olfr1264)	4	59,9	11,5
529	OR4C5	Olfactory receptor 4C5 (Olfr1260)	2	52	20,6
530	OR4A47	Olfactory receptor 4A47 (Olfr1256)	3	64	2,5

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
531	OR4C13	Olfactory receptor 4C13 (Olfr1258)	1	62,9	3,8
532	OR4C12	Olfactory receptor 4C12 (Olfr1255)	2	65,4	3,5
533	OR4C12	Olfactory receptor 4C12 (Olfr1259)	8	68,3	0,6
534	OR4C45	Olfactory receptor 4C45 (Olfr1262)	2	55,9	8,6
535	OR4C45	Olfactory receptor 4C45 (Olfr1263)	2	56,2	10,1
536	OR4C45	Olfactory receptor 4C45 (Olfr1261)	1	59,1	6,1
537	OR4C46	Olfactory receptor 4C46 (Olfr1258)	2	62,2	3,8
538	OR4A15	Olfactory receptor 4A15 (Olfr1234)	1	57,3	10
539	OR4C15	Olfactory receptor 4C15 (Olfr1212)	1	49,7	21,6
540	OR4C15	Olfactory receptor 4C15 (Olfr1218)	0	45,1	25,8
541	OR4C15	Olfactory receptor 4C15 (Olfr1232)	1	46,1	23,2
542	OR4C15	Olfactory receptor 4C15 (Olfr1216)	1	43,1	23,6
543	OR4C15	Olfactory receptor 4C15 (Olfr1229)	1	49,7	22,8
544	OR4C15	Olfactory receptor 4C15 (Olfr1214)	1	45	26,3
545	OR4C15	Olfactory receptor 4C15 (Olfr1226)	1	49,6	22,1
546	OR4C15	Olfactory receptor 4C15 (Olfr1223)	1	47,2	26,5
547	OR4C15	Olfactory receptor 4C15 (Olfr1211)	1	53,2	18,2
548	OR4C15	Olfactory receptor 4C15 (Olfr1228)	1	49,5	20,6
549	OR4C15	Olfactory receptor 4C15 (Olfr1221)	0	45,9	22,8
550	OR4C15	Olfactory receptor 4C15 (Olfr1215)	0	46,7	27,9
551	OR4C15	Olfactory receptor 4C15 (Olfr1219)	1	48	24,9
552	OR4C15	Olfactory receptor 4C15 (Olfr1217)	0	46,9	23,8
553	OR4C15	Olfactory receptor 4C15 (Olfr1220)	1	44,8	24,2
554	OR4C15	Olfactory receptor 4C15 (Olfr1225)	0	48,3	21,9
555	OR4C16	Olfactory receptor 4C16 (Olfr1209)	5	61,6	5,6
556	OR4C16	Olfactory receptor 4C16 (Olfr1195)	4	55	9,4
557	OR4C16	Olfactory receptor 4C16 (Olfr1199)	4	58,8	2,3
558	OR4C11	Olfactory receptor 4C11 (Olfr1205)	5	63,2	5,4
559	OR4C11	Olfactory receptor 4C11 (Olfr1201)	5	63,7	3,3
560	OR4C11	Olfactory receptor 4C11 (Olfr1206)	5	66,2	3,3
561	OR4S2	Olfactory receptor 4S2 (Olfr1193)	4	63,5	2,8
562	OR4C6	Olfactory receptor 4C6 (Olfr1189)	0	52,3	9,6
563	OR4C6	Olfactory receptor 4C6 (Olfr1230)	3	54,8	8,7
564	OR4D10	Olfactory receptor 4D10 (Olfr1426)	4	66	5,6
565	OR4D10	Olfactory receptor 4D10 (Olfr1425)	5	68,9	2,3
566	OR4D10	Olfactory receptor 4D10 (Olfr1424)	2	65,6	3,6
567	OR4D5	Olfactory receptor 4D5 (Olfr984)	2	66,3	3,9
568	OR4Q3	Olfactory receptor 4Q3 (Olfr735)	1	53,8	16,6
569	OR4M1	Olfactory receptor 4M1 (Olfr734)	4	67,2	1,3
570	OR4K2	Olfactory receptor 4K2 (Olfr730)	2	64	6,5
571	OR4K5	Olfactory receptor 4K5 (Olfr729)	4	63,2	2,8
572	OR4K15	Olfactory receptor 4K15 (Olfr727)	3	64,5	10,9
573	OR4K15	Olfactory receptor 4K15 (Olfr726)	2	60,8	11,2
574	OR4K15	Olfactory receptor 4K15 (Olfr725)	4	59,9	12,2
575	OR4L1	Olfactory receptor 4L1 (Olfr723)	4	61,6	3,3
576	OR4L1	Olfactory receptor 4L1 (Olfr724)	3	61	3,5
577	OR4E2	Olfactory receptor 4E2 (Olfr1509)	6	68	3,7
578	OR4M2	Olfactory receptor 4M2 (Olfr734)	4	64,8	1,3
579	OR4N4	Olfactory receptor 4N4 (Olfr733)	1	57,4	5,4
580	OR4N4	Olfactory receptor 4N4 (Olfr732)	1	57,9	8,6
581	OR4F15	Olfactory receptor 4F15 (Olfr1309)	2	63,5	4,7
582	OR4F4	Olfactory receptor 4F4 (Olfr1289)	3	59,5	3,8
583	OR1S1	Olfactory receptor 1S1 (Olfr1496)	0	28,9	35
584	OR1A2	Olfactory receptor 1A2 (Olfr43)	2	58,6	4
585	OR1A2	Olfactory receptor 1A2 (Olfr403)	3	58,6	4,2
586	OR1A1	Olfactory receptor 1A1 (Olfr43)	3	63,8	3,8
587	OR1A1	Olfactory receptor 1A1 (Olfr403)	3	63,4	5,6
588	OR8G1	Olfactory receptor 8G1 (Olfr27)	2	53,1	5,8
589	OR8G1	Olfactory receptor 8G1 (Olfr948)	2	51,4	11,5
590	OR8G1	Olfactory receptor 8G1 (Olfr44)	1	59,1	8,3
591	OR8G1	Olfactory receptor 8G1 (Olfr945)	1	52,7	12,7
592	OR8G1	Olfactory receptor 8G1 (Olfr937)	1	57,9	7,4
593	OR8G1	Olfactory receptor 8G1 (Olfr951)	1	53,2	11
594	OR8G1	Olfactory receptor 8G1 (Olfr954)	2	53,9	7,5
595	OR8G1	Olfactory receptor 8G1 (Olfr943)	0	52,3	10,4
596	OR8G1	Olfactory receptor 8G1 (Olfr1537)	1	54	7,1
597	OR8G1	Olfactory receptor 8G1 (Olfr944)	1	53,1	7,7

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
598	OR1J1	Olfactory receptor 1J1 (Olfr3)	3	61	8,1
599	OR6Q1	Olfactory receptor 6Q1 (Olfr522)	0	28,8	35,8
600	OR6C1	Olfactory receptor 6C1 (Olfr802)	5	65	0
601	OR6C1	Olfactory receptor 6C1 (Olfr786)	2	65,5	0
602	OR2L13	Olfactory receptor 2L13 (Olfr166)	3	57,5	3,6
603	OR2L13	Olfactory receptor 2L13 (Olfr168)	2	57,9	3,2
604	OR5B21	Olfactory receptor 5B21 (Olfr1444)	4	63,5	4,4
605	OR5D18	Olfactory receptor 5D18 (Olfr1161)	2	53,7	12
606	OR5D18	Olfactory receptor 5D18 (Olfr73)	2	58,4	5,6
607	OR5D18	Olfactory receptor 5D18 (Olfr74)	2	58,2	7,6
608	OR51I2	Olfactory receptor 51I2 (Olfr641)	4	73,2	0
609	OR52D1	Olfactory receptor 52D1 (Olfr557)	0	31,8	38,2
610	OR1J4	Olfactory receptor 1J4 (Olfr350)	1	66,4	3,9
611	OR1N1	Olfactory receptor 1N1 (Olfr353)	3	61,4	4,8
612	OR1N1	Olfactory receptor 1N1 (Olfr351)	1	62,8	3,7
613	OR2A2	Olfactory receptor 2A2 (Olfr437)	1	55,9	10,5
614	OR10K2	Olfactory receptor 10K2 (Olfr370)	4	58,8	5,8
615	OR2T4	Olfactory receptor 2T4 (Olfr328)	2	55,7	13,8
616	OR2T4	Olfactory receptor 2T4 (Olfr331)	2	57,3	12,8
617	OR2T4	Olfactory receptor 2T4 (Olfr224)	2	55,1	13,9
618	OR2T4	Olfactory receptor 2T4 (Olfr325)	2	56,1	14,7
619	OR10A5	Olfactory receptor 10A5 (Olfr713)	7	74,3	2,2
620	OR10A5	Olfactory receptor 10A5 (Olfr714)	5	72,2	0
621	OR6N1	Olfactory receptor 6N1 (Olfr429)	2	68,3	1,9
622	OR9G1	Olfactory receptor 9G1 (Olfr1014)	4	57,7	7
623	OR9G1	Olfactory receptor 9G1 (Olfr1016)	4	54,9	8,7
624	OR1N2	Olfactory receptor 1N2 (Olfr354)	4	66,5	5,7
625	OR2J1	Olfactory receptor 2J1 (Olfr137)	3	57,1	3,8
626	OR6C74	Olfactory receptor 6C74 (Olfr821)	3	70,8	1,5
627	OR9I1	Olfactory receptor 9I1 (Olfr1499)	1	54,9	11
628	OR9I1	Olfactory receptor 9I1 (Olfr1502)	1	56,6	6,1
629	OR9I1	Olfactory receptor 9I1 (Olfr1505)	1	56,2	7,5
630	OR7G2	Olfactory receptor 7G2 (Olfr855)	1	47,2	14,7
631	OR7G2	Olfactory receptor 7G2 (Olfr830)	2	44,1	19,2
632	OR7G2	Olfactory receptor 7G2 (Olfr835)	0	43,1	18,6
633	OR7G2	Olfactory receptor 7G2 (Olfr828)	1	44,8	17,8
634	OR7G2	Olfactory receptor 7G2 (Olfr843)	1	41,8	15,8
635	OR7G2	Olfactory receptor 7G2 (Olfr850)	3	47,8	19,2
636	OR7G2	Olfactory receptor 7G2 (Olfr829)	1	44,5	21,1
637	OR7G2	Olfactory receptor 7G2 (Olfr854)	3	50,2	18,3
638	OR7G2	Olfactory receptor 7G2 (Olfr849)	2	43,3	14,9
639	OR7G2	Olfactory receptor 7G2 (Olfr837)	2	42,7	20,1
640	OR7G2	Olfactory receptor 7G2 (Olfr851)	0	41,2	20,3
641	OR7G2	Olfactory receptor 7G2 (Olfr846)	3	43,3	18,2
642	OR7G2	Olfactory receptor 7G2 (Olfr836)	2	44,8	20,3
643	OR7G2	Olfactory receptor 7G2 (Olfr847)	4	42,8	19,5
644	OR7G2	Olfactory receptor 7G2 (Olfr845)	1	46,6	15,4
645	OR7A17	Olfactory receptor 7A17 (Olfr1353)	1	53,2	10,4
646	OR7A17	Olfactory receptor 7A17 (Olfr1352)	0	51,6	9,6
647	OR7A17	Olfactory receptor 7A17 (Olfr19)	1	56,4	2,3
648	OR7A17	Olfactory receptor 7A17 (Olfr57)	1	57,7	14,4
649	OR7A17	Olfactory receptor 7A17 (Olfr1355)	0	46,8	16,5
650	OR7A17	Olfactory receptor 7A17 (Olfr1351)	0	55,1	15,8
651	OR7A17	Olfactory receptor 7A17 (Olfr8)	1	45,2	21,2
652	OR7A17	Olfactory receptor 7A17 (Olfr1354)	0	45,8	20,9
653	OR8D1	Olfactory receptor 8D1 (Olfr26)	3	64,7	5,3
654	OR8D1	Olfactory receptor 8D1 (Olfr933)	2	63	5,7
655	OR8D1	Olfactory receptor 8D1 (Olfr930)	2	62,4	4
656	OR2AG1	Olfactory receptor 2AG1 (Olfr701)	1	62,5	3,3
657	OR2AG1	Olfactory receptor 2AG1 (Olfr706)	3	61	5,9
658	OR6Y1	Olfactory receptor 6Y1 A5 (Olfr220)	5	64,2	3,6
659	OR1S2	Olfactory receptor 1S2 (Olfr1496)	2	59,3	8,2
660	OR6C76	Olfactory receptor 6C76 (Olfr809)	3	62,1	11
661	OR6C76	Olfactory receptor 6C76 (Olfr813)	4	65,5	4,2
662	OR8K5	Olfactory receptor 8K5 (Olfr1008)	1	55,6	7
663	OR8K5	Olfactory receptor 8K5 (Olfr1053)	2	58,6	6
664	OR8K5	Olfactory receptor 8K5 (Olfr1051)	1	57,4	1,8

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
665	OR8K5	Olfactory receptor 8K5 (Olfr1049)	0	56,8	2
666	OR8K5	Olfactory receptor 8K5 (Olfr1048)	3	52	9,8
667	OR1L6	Olfactory receptor 1L6 (Olfr365)	1	64,3	4,5
668	OR2F1	Olfactory receptor 2F1 (Olfr452)	7	64	6,7
669	OR2F1	Olfactory receptor 2F1 (Olfr38)	8	66,7	5,5
670	OR2F1	Olfactory receptor 2F1 (Olfr453)	5	62,7	2,7
671	OR4N5	Olfactory receptor 4N5 (Olfr722)	0	29,5	41,2
672	OR4D6	Olfactory receptor 4D6 (Olfr1428)	5	65,4	1,9
673	OR1L4	Olfactory receptor 1L4 (Olfr365)	2	65,3	5,5
674	OR6X1	Olfactory receptor 6X1 (Olfr986)	5	66,9	1,7
675	OR52I1	Olfactory receptor 52I1 (Olfr640)	0	27	42,4
676	OR51T1	Olfactory receptor 51T1 (Olfr574)	0	54,4	14,5
677	OR9Q1	Olfactory receptor 9Q1 (Olfr1500)	4	65	0,7
678	OR52B4	Olfactory receptor 52B4 (Olfr547)	6	66,9	3
679	OR52I2	Olfactory receptor 52I2 (Olfr556)	0	55,9	10,8
680	OR6C3	Olfactory receptor 6C3 (Olfr788)	4	68,8	3
681	OR6C6	Olfactory receptor 6C6 (Olfr779)	2	64,8	3,1
682	OR6C6	Olfactory receptor 6C6 (Olfr782)	2	65,8	0,8
683	OR6C6	Olfactory receptor 6C6 (Olfr804)	2	66,2	0
684	OR52H1	Olfactory receptor 52H1 (Olfr648)	4	65,5	7,7
685	OR52L1	Olfactory receptor 52L1 (Olfr685)	2	63,6	5,1
686	OR7G3	Olfactory receptor 7G3 (Olfr834)	2	51,1	5,6
687	OR5AS1	Olfactory receptor 5AS1 (Olfr1111)	2	56,3	10,3
688	OR5AC2	Olfactory receptor 5AC2 (Olfr202)	3	52,5	9,7
689	OR5AC2	Olfactory receptor 5AC2 (Olfr198)	2	49,2	15,9
690	OR5AC2	Olfactory receptor 5AC2 (Olfr201)	2	47,1	19,6
691	OR5AC2	Olfactory receptor 5AC2 (Olfr199)	3	47,2	17,1
692	OR2A12	Olfactory receptor 2A12 (Olfr446)	4	58,5	7,5
693	OR7A5	Olfactory receptor 7A5 (Olfr1352)	1	46,4	21,4
694	OR7A5	Olfactory receptor 7A5 (Olfr19)	0	53,9	9,7
695	OR7A5	Olfactory receptor 7A5 (Olfr57)	1	54,2	14
696	OR7A5	Olfactory receptor 7A5 (Olfr1355)	1	46,9	20,3
697	OR7A5	Olfactory receptor 7A5 (Olfr1351)	3	54,6	12
698	OR7A5	Olfactory receptor 7A5 (Olfr8)	1	46,9	19,4
699	OR7A5	Olfactory receptor 7A5 (Olfr1354)	1	46,9	19,1
700	OR5M1	Olfactory receptor 5M1 (Olfr1023)	4	62,9	5,6
701	OR5M1	Olfactory receptor 5M1 (Olfr1022)	6	63,2	1,9
702	OR8H2	Olfactory receptor 8H2 (Olfr1100)	2	54,1	11,6
703	OR8H2	Olfactory receptor 8H2 (Olfr1099)	1	52,5	8
704	OR8H2	Olfactory receptor 8H2 (Olfr1098)	2	53,5	9,8
705	OR8H2	Olfactory receptor 8H2 (Olfr1097)	2	53,9	7,9
706	OR5D14	Olfactory receptor 5D14 (Olfr1163)	2	52,8	8,2
707	OR5D14	Olfactory receptor 5D14 (Olfr1162)	3	54,8	6,8
708	OR5D14	Olfactory receptor 5D14 (Olfr868)	0	31,1	32,5
709	OR7E24	Olfactory receptor 7E24 (Olfr873)	2	53,2	13,1
710	OR7E24	Olfactory receptor 7E24 (Olfr866)	3	47,4	14,5
711	OR7E24	Olfactory receptor 7E24 (Olfr869)	1	50,3	12,6
712	OR7E24	Olfactory receptor 7E24 (Olfr872)	2	49,3	15,8
713	OR52J3	Olfactory receptor 52J3 (Olfr592)	3	61,2	4,5
714	OXTR	Oxytocin receptor	6	69,2	9,2
715	P2RY1	P2Y purinoceptor 1	7	69,7	1,1
716	P2RY10	Putative P2Y purinoceptor 10	1	47,2	23
717	P2RY12	P2Y purinoceptor 12	2	60,9	5,2
718	P2RY4	P2Y purinoceptor 4	2	63,7	6,4
719	P2RY5	Lysophosphatidic acid receptor 6	3	70,2	1,7
720	P2RY6	P2Y purinoceptor 6	1	63	5
721	PPYR1	Neuropeptide Y receptor type 4	0	56,3	9,9
722	PTAFR	Platelet-activating factor receptor	0	57,9	13,6
723	PTGDR	Prostaglandin D2 receptor 2	3	55,6	11,9
724	PTGER1	Prostaglandin E2 receptor EP1 subtype	2	58,9	8,3
725	PTGER2	Prostaglandin E2 receptor EP2 subtype	8	65,3	2,6
726	PTGER3	Prostaglandin E2 receptor EP3 subtype	6	58,6	16,5
727	PTGER4	Prostaglandin E2 receptor EP4 subtype	5	68,7	7,4
728	PTGFR	Prostaglandin F2-alpha receptor	2	49,7	21,9
729	PTGIR	Prostacyclin receptor	4	55,6	16,5
730	PTHR1	Parathyroid hormone/parathyroid	6	69,3	5,5
731	PTHR2	Parathyroid hormone 2 receptor	6	64	4,9

Nº	Gene name	Description	Shared tCPs	tCP identity (%)	Gaps (%)
732	RRH	Visual pigment-like receptor peropsin	2	61,4	5,8
733	RLN3R1	Relaxin-3 receptor 1	5	65,7	4,1
734	SCTR	Secretin receptor	5	64,5	8
735	SMO	Smoothened homolog	5	75,6	31,7
736	SSTR1	Somatostatin receptor type 1	7	73,1	3,7
737	SSTR2	Somatostatin receptor type 2	5	71,6	5,6
738	SSTR3	Somatostatin receptor type 3	5	61,7	11,3
739	SSTR4	Somatostatin receptor type 4	4	65	7,5
740	SSTR5	Somatostatin receptor type 5	1	50,4	19,9
741	TACR1	Substance-P receptor	6	75,9	1,8
742	TACR2	Substance-K receptor	2	61,8	11
743	TACR3	Neuromedin-K receptor	6	67,4	7,2
744	TRHR	Thyrotropin-releasing hormone receptor	6	76,1	2,6
745	TSHR	Thyrotropin receptor	6	68,2	3,3

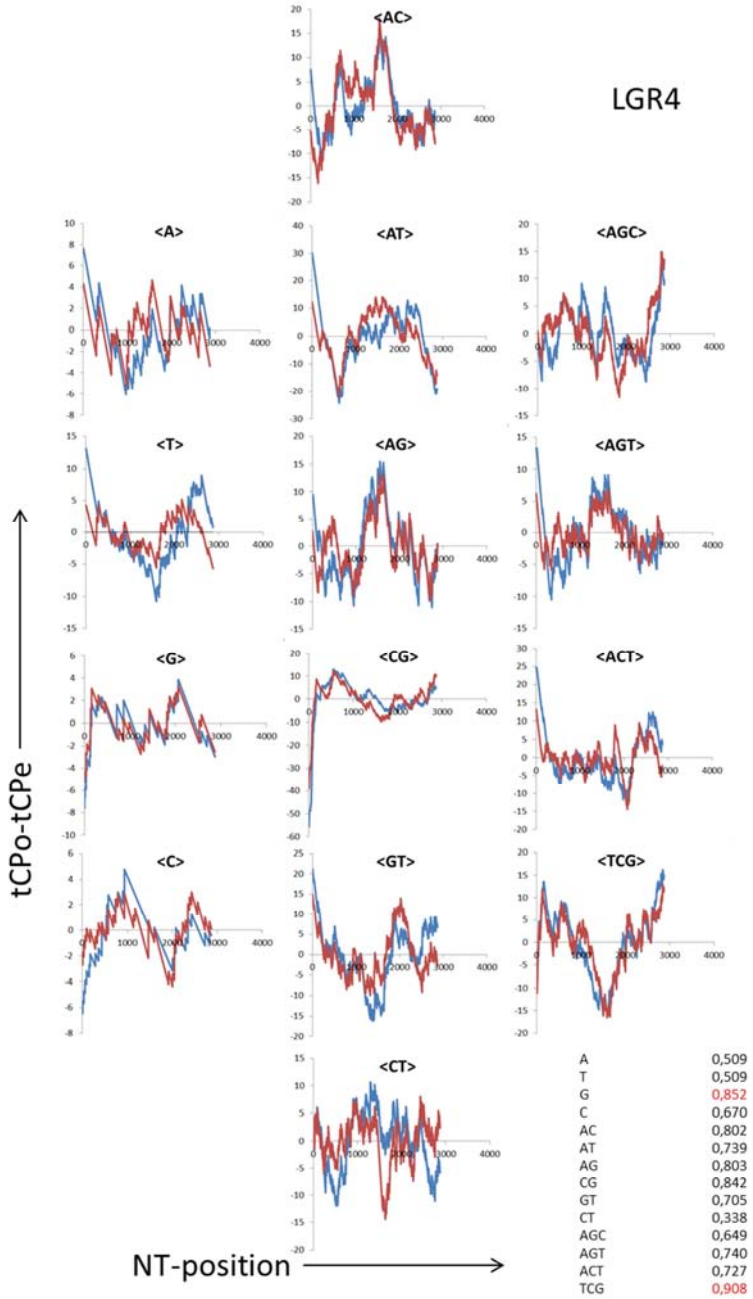

Figure 8. *Comparison of tCP-profiles from the human-mouse ortholog LGR4. Red and blue lines correspond to the profiles of mouse and human, respectively. The insets display the name of the ortholog (upper right corner), and a table with the Pearson correlation coefficient, r, between tCP-profiles of human and mouse (lower right corner). Pearson correlation coefficient for r≥0.85 are in red.*

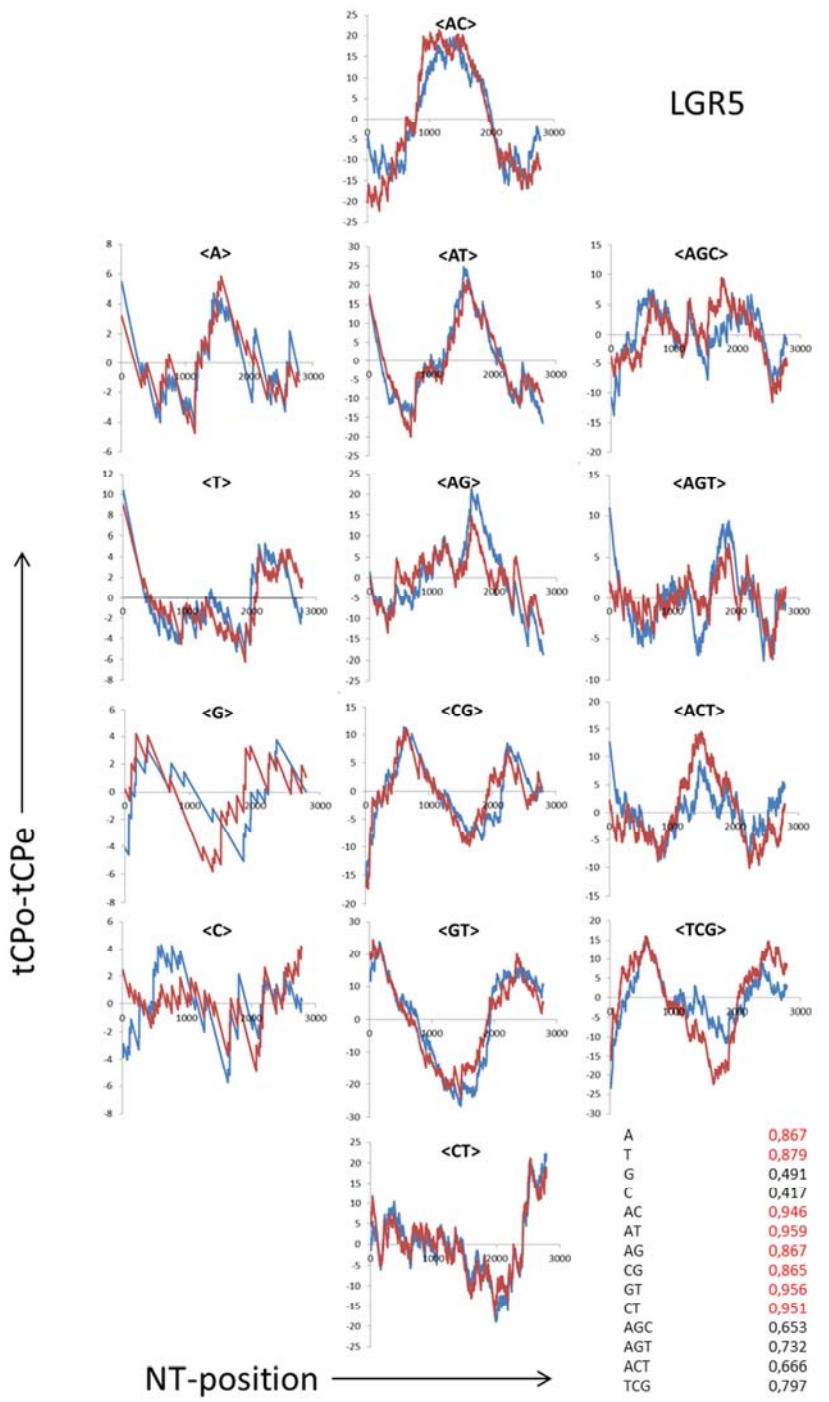

Figure 9. *Comparison of tCP-profiles from the human-mouse ortholog LGR4. Red and blue lines correspond to the profiles of mouse and human, respectively. The insets display the name of the ortholog (upper right corner), and a table with the Pearson correlation coefficient, r, between tCP-profiles of human and mouse (lower right corner). Pearson correlation coefficient for r≥0.85 are in red.*

Data Accessibility Statement

Materials, data and associated protocols are promptly available to readers without undue qualification in material transfer agreement.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgements

I would like to express my deep gratitude to Prof Carlos Alonso Bedate who died on April 13, 2020 in the middle of the SARS CoV-2 pandemic for his advice, guidance, friendship and encouragement throughout the last twenty-five years of my life. An institutional grant from Foundation Ramón Areces is also acknowledged.

References

[1] R. Fredriksson, M. C. Lagerstrom, L. G. Lundin, H. B. Schioth, The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63, 1256-1272 (2003).

[2] D. M. Rosenbaum, S. G. Rasmussen, B. K. Kobilka, The structure and function of G-protein-coupled receptors. Nature 459, 356-363 (2009).

[3] M. Rask-Andersen, M. S. Almén, H. B. Schiöth. Trends in the exploitation of novel drug targets. Nat Rev Drug Discov 10, 579–590 (2011).

[4] O. S. Soyer, M. W. Dimmic, R. R. Neubig, R. A. Goldstein. Dimerization in aminergic G-protein-coupled receptors: application of a hidden-site class model of evolution. Biochemistry 42,14522-14531 (2003).

[5] W. K. Kroeze, D. J. Sheffler, B. L. Roth, G-protein-coupled receptors at a glance. J Cell Sci 116, 4867-4869 (2003).

[6] Y. Zhang, M. E. Devries, J. Skolnick, Structure modeling of all identified G protein-coupled receptors in the human genome. PLoS Comput Biol 2, e13 (2006).

[7] E. van der Horst, J. E. Peironcely, A. P. Ijzerman, M. W. Beukers, J. R. Lane, H. W. van Vlijmen, M. T. Emmerich, Y. Okuno, A. Bender, A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: a potential strategy for receptor de-orphanization. BMC Bioinformatics 11, 316 (2010).

[8] E. Jacoby, R. Bouhelal, M. Gerspacher, K. Seuwen. The 7 TM G-protein-coupled receptor target family. ChemMedChem 1, 761-782 (2006).

[9] Y. Kawasawa, L. M. McKenzie, D. P. Hill, H. Bono, M. Yanagisawa, RIKEN GER Group; GSL Members, G protein-coupled receptor genes in the FANTOM2 database. Genome Res 13, 1466-1477 (2003).

[10] M. Wistrand, L. Kall, E. L. Sonnhammer. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. Protein Sci 15, 509-521 (1991).

[11] Buck. L, Axel. R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell 65, 175-187 (2006).

[12] Y. Pilpel, D. Lancet. The variable and conserved interfaces of modeled olfactory receptor proteins. Protein Sci 8, 969-977 (1999).

[13] W. C. Probst, L. A. Snyder, D. I. Schuster, J. Brosius, S. C. Sealfon. Sequence alignment of the G-protein coupled receptor superfamily. DNA Cell Biol 11, 1-20 (1992).

[14] R. A. Gibbs, G. M. Weinstock, M. L. Metzker, D. M. Muzny, E. J. Sodergren, S. Scherer, G. Scott, D. Steffen, K. C. Worley, P. E. Burch, G. Okwuonu, S. Hines, L. Lewis, C. DeRamo, O. Delgado, S. Dugan-Rocha, G. Miner, M. Morgan, A. Hawes, R. Gill, Celera, R. A. Holt, M. D. Adams, P. G. Amanatides, H. Baden-Tillson, M. Barnstead, S. Chin, C. A. Evans, S. Ferriera, C. Fosler, A. Glodek, Z. Gu, D. Jennings, C. L. Kraft, T. Nguyen, C. M. Pfannkoch, C. Sitter, G. G. Sutton, J. C. Venter, T. Woodage, D. Smith, H. M. Lee, E. Gustafson, P. Cahill, A. Kana, L. Doucette-Stamm, K. Weinstock, K. Fechtel, R. B. Weiss, D. M. Dunn, E. D. Green, R. W. Blakesley, G. G. Bouffard, P. J. De Jong, K. Osoegawa, B. Zhu, M. Marra, J. Schein, I. Bosdet, C. Fjell, S. Jones, M. Krzywinski, C. Mathewson, A. Siddiqui, N. Wye, J. McPherson, S. Zhao, C. M. Fraser, J. Shetty, S. Shatsman, K. Geer, Y. Chen, S. Abramzon, W. C. Nierman, P. H. Havlak, R. Chen, K. J. Durbin, A. Egan, Y. Ren, X. Z. Song, B. Li, Y. Liu, X. Qin, S. Cawley, K. C. Worley, A. J. Cooney, L. M. D'Souza, K. Martin, J. Q. Wu, M. L. Gonzalez-Garay, A. R. Jackson, K. J. Kalafus, M. P. McLeod, A. Milosavljevic, D. Virk, A. Volkov, D. A. Wheeler, Z. Zhang, J. A. Bailey, E. E. Eichler, E. Tuzun, E. Birney, E. Mongin, A. Ureta-Vidal, C. Woodwark, E. Zdobnov, P. Bork, M. Suyama, D. Torrents, M. Alexandersson, B. J. Trask, J. M. Young, H. Huang, H. Wang, H. Xing, S. Daniels, D. Gietzen, J. Schmidt, K. Stevens, U. Vitt, J. Wingrove, F. Camara, M. Mar Alba, J. F. Abril, R. Guigo, A. Smit, I. Dubchak, E. M. Rubin, O. Couronne, A. Poliakov, N. Hubner, D. Ganten, C. Goesele, O. Hummel, T. Kreitler, Y. A. Lee, J. Monti, H. Schulz, H. Zimdahl, H. Himmelbauer, H. Lehrach, H. J. Jacob, S. Bromberg, J. Gullings-Handley, M. I. Jensen-Seaman, A. E. Kwitek, J. Lazar, D. Pasko, P. J. Tonellato, S. Twigger, C. P. Ponting, J. M. Duarte, S. Rice, L. Goodstadt, S. A. Beatson, R. D. Emes, E. E. Winter, C. Webber, P. Brandt, G. Nyakatura, M. Adetobi, F. Chiaromonte, L. Elnitski, P. Eswara, R. C. Hardison, M. Hou, D. Kolbe, K. Makova, W. Miller, A. Nekrutenko, C. Riemer, S. Schwartz, J. Taylor, S. Yang, Y. Zhang, K. Lindpaintner, T. D. Andrews, M. Caccamo, M. Clamp, L. Clarke, V. Curwen, R. Durbin, E. Eyras, S. M. Searle, G. M. Cooper, S. Batzoglou, M. Brudno, A. Sidow, E. A. Stone, J. C. Venter, B. A. Payseur, G. Bourque, C. Lopez-Otin, X. S. Puente, K. Chakrabarti, S. Chatterji, C. Dewey, L. Pachter, N. Bray, V. B. Yap, A. Caspi, G. Tesler, P. A. Pevzner, D. Haussler, K. M. Roskin, R. Baertsch, H. Clawson, T. S. Furey, A. S. Hinrichs, D. Karolchik, W. J. Kent, K. R. Rosenbloom, H. Trumbower, M. Weirauch, D. N. Cooper, P. D. Stenson, B. Ma, M. Brent, M. Arumugam, D. Shteynberg, R. R. Copley, M. S. Taylor, H. Riethman, U. Mudunuri, J. Peterson, M. Guyer, A. Felsenfeld, S. Old, S. Mockrin, F. Collins. Rat Genome Sequencing Project C Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428, 493-521 (2004).

[15] M. Behrens, L. Briand, C. A. de March, H. Matsunami, A. Yamashita, W. Meyerhof, S. Weyand. Structure-Function Relationships of Olfactory and Taste Receptors. Chem Senses 43, 81-87 (2018).

[16] T. Warne, R. Moukhametzianov, J. G. Baker, R. Nehme, P. C. Edwards, A. G. Leslie, G. F. Schertler, C. G. Tate. The structural basis for agonist and partial agonist action on a beta (1)-adrenergic receptor. Nature 469, 241-244 (2011).

[17] L. Charlier, J. Topin, C. A. de March, P. C. Lai, C. J. Crasto, J. Golebiowski. Molecular modelling of odorant/olfactory receptor complexes. Methods Mol Biol 1003, 53-65 (2013).

[18] P. C. Lai, B. Guida, J. Shi, C. J. Crasto. Preferential binding of an odor within olfactory receptors: a precursor to receptor activation. Chem Senses 39, 107-123 (2014).

[19] P. C. Lai, M. S. Singer, C. J. Crasto. Structural activation pathways from dynamic olfactory receptor-odorant interactions. Chem Senses 30, 781-792 (2005).

[20] M. A. Fuertes, J. M. Pèrez, E. Zuckerkandl, C. Alonso Introns form compositional clusters in parallel with the compositional clusters of the coding sequences to which they pertain. J Mol Evol 72, 1-13 (2011).

[21] M. A. Fuertes, C. Alonso. A method for the Annotation of Functional Similarities of Coding DNA Sequences: the Case of a Populated Cluster of Transmembrane Proteins. J Mol Evol 84, 29-38 (2016).

[22] M. A. Fuertes, J. R. Rodrigo, C. Alonso. Conserved Critical Evolutionary Gene Structures in Orthologs. J Mol Evol 87, 93-105, (2019).

[23] M. A. Fuertes, S. Lòpez-Arguello, C. Alonso, C. Evolutionary conserved compositional structures hidden in genomes of the foot-and-mouth disease virus and of the human rhinovirus. *Sci Rep* 9, 16553 (2019).

[24] Y. M. Park, S. Squizzato, N. Buso, T. Gur, R. Lopez R. The EBI search engine: EBI search as a service-making biological data accessible for all. Nucleic Acids Res 45, W545-W549 (2017).

[25] M. Safran, V. Chalifa-Caspi, O. Shmueli, T. Olender, M. Lapidot, N. Rosen, M. Shmoish, Y. Peter, G. Glusman, E. Feldmesser, A. Adato, I. Peter, M. Khen, T. Atarot, Y. Groner, D. Lancet D. Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE. Nucleic Acids Res 31, 142-146 (2003).

[26] G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, . Kohn, N. Rappaport, M. Safran, D. Lancet. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics 54, 1.30.1-1.30.33 (2016).

[27] J. M. Young, B. J. Trask. The sense of smell: genomics of vertebrate odorant receptors. Hum Mol Genet 11, 1153-1160 (2002).

[28] D. D. Pervouchine, S. Djebali, A. Breschi, C. A. Davis, P. P. Barja, A. Dobin, A. Tanzer, J. Lagarde, C. Zaleski, L. H. See, M. Fastuca, J. Drenkow, H. Wang, G. Bussotti, B. Pei, S. Balasubramanian, J. Monlong, A. Harmanci, M. Gerstein, M. A. Beer, C. Notredame, R. Guigo, T. R. Gingeras. Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. Nat Commun 6, 5903 (2015).

[29] G. Glusman, A. Bahar, D. Sharon, Y. Pilpel, J. White, D. Lancet. The olfactory receptor gene superfamily: data mining, classification, and nomenclature. Mamm Genome 11, 1016-1023 (2000).

[30] C. L. Smith, J. A. Blake, J. A. Kadin, J. E. Richardson, C. J. Bult, G. Mouse Genome Database. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. Nucleic Acids Res 46, D836-D842 (2018).

[31] V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, R. C. Stevens. High-resolution crystal structure of an engineered human beta 2-adrenergic G protein-coupled receptor. Science 318, 1258-1265 (2007).

[32] F. Z. Chung, C. D. Wang, P. C. Potter, J. C. Venter, C. M. Fraser. Site-directed mutagenesis and continuous expression of human beta-adrenergic receptors. Identification of a conserved aspartate residue involved in agonist binding and receptor activation. J Biol Chem 263, 4052-4055 (1988).

[33] D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, X. J. Yao, W. I. Weis, R. C. Stevens, B. K. Kobilka. GPCR engineering yields high-resolution structural insights into beta 2-adrenergic receptor function. Science 318, 1266-1273 (2007).

[34] W. R. Pearson. An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinformatics Chapter 3, Unit 3.1 (2013).

[35] J. B. Kruskal. An overview of squence comparison. Time warps, string edits and macromolecules: the theory and practice of sequence comparison SIAM Review. CSLI Publications, Stanford University, pp. 201-237 (1983).

[36] S. B. Needleman, C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48, 443-453 (1970).

[37] L. J. Mullan, A. J. Bleasby. Short EMBOSS User Guide. European Molecular Biology Open Software Suite. Brief Bioinform 3, 92-94 (2002).

[38] S. A. Olson. EMBOSS opens up sequence analysis. European Molecular Biology Open Software Suite. Brief Bioinform 3, 87-91 (2002).

[39] M. Grossmann, B. D. Weintraub, M. W. Szkudlinski. Novel insights into the molecular mechanisms of human thyrotropin action: structural, physiological, and therapeutic implications for the glycoprotein hormone family. Endocr Rev 18, 476-501 (1997).

[40] A. P. N. Themmen, I. T. Huhtaniemi. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. Endocr Rev 21, 551-583 (2000).

[41] N. Barker, H. Clevers. Leucine-rich repeat-containing G-protein-coupled receptors as markers of adult stem cells. Gastroenterology 138, 1681-1696 (2010).

[42] S. Nakata, E. Phillips, V. Goidts. Emerging role for leucine-rich repeat-containing G-protein-coupled receptors LGR5 and LGR4 in cancer stem cells. Cancer Manag Res 6, 171-180 (2014).

[43] C. A. de March, Y. Yu, M. J. Ni, K. A. Adipietro, H. Matsunami, M. Ma, J. Golebiowski. Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. J Am Chem Soc 137, 8611-8616 (2015).

[44] A. S. Dore, K. Okrasa, J. C. Patel, M. Serrano-Vega, K. Bennett, R. M. Cooke, J. C. Errey, A. Jazayeri, S. Khan, B. Tehan, M. Weir, G. R. Wiggin, F. H. Marshall. Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. Nature 511, 557-562 (2014).

[45] E. Hodge, C. P. Nelson, S. Miller, C. K. Billington, C. E. Stewart, C. Swan, A. Malarstig, A. P. Henry, C. Gowland, E. Melén, I. P. Hall, I. Sayers. HTR4 gene structure and altered expression in the developing lung. Respir Res 14, 77 (2013).

[46] T. K. Bjarnadottir, D. E. Gloriam, S. H. Hellstrand, H. Kristiansson, R. Fredriksson, H. B. Schioth. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. Genomics 88, 263-273 (2006).

[47] M. Johnson. Molecular mechanisms of beta (2)-adrenergic receptor function, response, and regulation. J Allergy Clin Immunol 117, 18-24 (2006).

[48] H. C. Huang, P. S. Klein. The Frizzled family: receptors for multiple signal transduction pathways. Genome Biol 5, 234 (2004).

[49] H. Clevers, K. M. Loh, R. Nusse. Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. Science 346, 1248012 (2014).

[50] S. Y. Hsu. New insights into the evolution of the relaxin-LGR signaling system. Trends Endocrinol Metab 14, 303-309 (2003).

[51] W. de Lau, W. C. Peng, P. Gros, H. Clevers. The R-spondin/Lgr5/Rnf43 module: regulator of Wnt signal strength. Genes Dev 28, 305-316 (2014).

[52] K. S. Carmon, X. Gong, Q. Lin, A. Thomas, Q. Liu. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/beta-catenin signaling. Proc Natl Acad Sci U S A 108, 11452-11457 (2011).

[53] R. Chen, E. V. Davydov, M. Sirota, A. J. Butte. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. PLoS One 5, e13574 (2010).

[54] R. Hunt, Z. E. Sauna, S. V. Ambudkar, M. M. Gottesman, C. Kimchi-Sarfaty. Silent (synonymous) SNPs: should we care about them? Methods Mol Biol 578, 23-39 (2009).

[55] S. A. Shabalina, N. A. Spiridonov, A. Kashina. Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. Nucleic Acids Res 41, 2073-2094 (2013).

[56] D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke, F. H. Arnold. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A 102 (40), 14338-14343, (2005).

[57] D. A. Drummond, A. Raval, C. O. Wilke. A single determinant dominates the rate of yeast protein evolution. *Mol Biol Evol* 23, 327-337 (2006).

[58] D. A. Drummond, C. O. Wilke. Mistranslation-induced protein misfolding as a dominantconstraint on coding-sequence evolution. *Cell* 134 (2), 341-352 (2008).

[59] D. Agashe, N. C. Martinez-Gomez, D. A. Drummond, C. J. Marx. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. Mol Biol Evol 30, 549-560 (2013).

[60] M. J. Lajoie, S. Kosuri, J. A. Mosberg, C. J. Gregg, D. Zhang, G. M. Church. Probing the limits of genetic recoding in essential genes. Science 342, 361-363 (2013a).

[61] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H. R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs. Genomically recoded organisms expand biological functions. Science 342, 357-360 (2013b).

[62] N. Ostrov, M. Landon, M. Guell, G. Kuznetsov, J. Teramoto, N. Cervantes, M. Zhou, K. Singh, M. G. Napolitano, M. Moosburner, E. Shrock, B. W. Pruitt, N. Conway, D. B. Goodman, C. L. Gardner, G. Tyree, A. Gonzales, B. L. Wanner, J. E. Norville, M. J. Lajoie, G. M. Church. Design, synthesis, and testing toward a 57-codon genome. Science 353, 819-822 (2016).

[63] J. H. Felce, S. L. Latty, R. G. Knox, S. R. Mattick, Y. Lui, S. F. Lee, D. Klenerman, S. J. Davis. Receptor Quaternary Organization Explains G Protein-Coupled Receptor Family Structure. Cell Rep 20, 2654-2665 (2017).

[64] N. Ben-Arie, D. Lancet, C. Taylor, M. Khen, N. Walker, D. H. Ledbetter, R. Carrozzo, K. Patel, D. Sheer, H. Lehrach, M. A. North. Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire. Hum Mol Genet 3, 229-235 (1994).

[65] G. Glusman, S. Clifton, B. Roe, D. Lancet. Sequence analysis in the olfactory receptor gene cluster on human chromosome 17: recombinatorial events affecting receptor diversity. Genomics 37, 147-160 (1996).

[66] M. C. Lagerstrom, H. B. Schioth. Structural diversity of G protein-coupled receptors and significance for drug discovery. Nat Rev Drug Discov 7, 339-357 (2008).

[67] M. Nakamura, D. Yasuda, N. Hirota, T. Shimizu. Specific ligands as pharmacological chaperones: The transport of misfolded G-protein coupled receptors to the cell surface. IUBMB Life 62, 453-459 (2010).

[68] C. E. Elling, K. Thirstrup, B. Holst, T. W. Schwartz. Conversion of agonist site to metal-ion chelator site in the beta (2)-adrenergic receptor. Proc Natl Acad Sci U S A 96, 12322-12327 (1999).

[69] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano. Crystal structure of hodopsin: A G protein-coupled receptor. Science 289, 739-745 (2000).