Vol.4, No.12A, 1109-1152 (2012)
http://dx.doi.org/10.4236/ns.2012.412A139

Discovery and validation of potential drug targets based on the phylogenetic evolution of GPCRs

Jie Yang*, Sen Li, Tongyang Zhu, Xiaoning Wang, Zhen Zhang

State Key Laboratory of Pharmaceutical Biotechnology, College of Life Sciences, Nanjing University, Nanjing, China;
*Corresponding Author: yangjie@nju.edu.cn

Received 8 October 2012; revised 10 November 2012; accepted 23 November 2012

ABSTRACT

Target identification is a critical step following the discovery of small molecules that elicit a biological phenotype. G-protein coupled recaptors (GPCRs) are among the most important drug targets for the pharmaceutical industry. The present work seeks to provide an *in silico* model of known GPCR protein fishing technologies in order to rapidly fish out potential drug targets on the basis of amino acid sequences and seven transmembrane regions (TMs) of GPCRs. Some scoring matrices were trained on 22 groups of GPCRs in the GPCRDB database. These models were employed to predict the GPCR proteins in two groups of test sets. On average, the mean correct rate of each TM of 38 GPCRs from two test sets (S_{23}^{T} and S_{24}^{T}) was found 62% and 57.5%, respectively, using training set 18 (S_{D18}^{L}); the mean hit rate of each TM of 38 GPCRs from S_{23}^{T} and S_{24}^{T} was found 68.1% and 64.7%, respectively. Based on the scoring matrices of PreMod, the mean correct rate of each TM of GPCRs from S_{23}^{T} and S_{24}^{T} was found 62% and 62.04%, respectively; the mean hit rate of each TM of GPCRs from S_{23}^{T} and S_{24}^{T} was found 67.7% and 68.0%, respectively. The means of GPCRs in S_{23}^{T} based on S_{D18}^{L} is close to those based on PreMod; whereas the means of GPCRs in S_{24}^{T} based on S_{D18}^{L} is less than those based on PreMod. Moreover, the accuracy ("2") and validity ("2 + 1") rates of prediction all seven TMs of 38 GPCRs by the scoring matrices of PreMod are more than those by S_{D18}^{L}, S_{A14}^{L} and S_{A3}^{L}; whereas the hit rates (94.74% and 97.37%) by PreMod are less than those of S_{A3}^{L} but bigger than those of S_{D18}^{L} and S_{A14}^{L}, respectively. This is the reason that we choose PreMod to predict some potential drug targets. 22 GPCR proteins in the sense chain of chromosome 19 constructing validation set were predicted and validated by PreMod whose hit rate is up to 90.91%. Further evaluation is under investigation.

Keywords: Pharmaceutical Targets for Drug Development; G-Protein Coupled Receptors; Scoring Matrices; Hit Rates

1. INTRODUCTION

G-protein coupled receptors (GPCRs) are among the most important drug targets for the pharmaceutical industry [1]. More than 30% of all marketed therapeutics interacts with them. GPCRs are integral membrane proteins that possess seven membrane-spanning domain or transmembrane helices with the N terminal of these proteins located in extracellular and the C-terminal extended in the cytoplasm. They comprise a large protein family of transmembrane receptors that sense molecules outside the cell and activate inside signal transduction pathways and, ultimately, cellular responses. The heterotrimeric G proteins (guanine nucleotide-binding proteins) are signal transducers, attached to the cell surface plasma membrane, that connect receptors to effectors and thus to intracellular signaling pathways [2,3]. The extracellular signals are received by GPCRs that activate the G proteins, which communicate signals from many hormones, neurotransmitters, chemokines, and autocrine and paracrine factors by several distinct intracellular signaling pathways [2]. These pathways interact with one another to form a network that regulates metabolic enzymes, ion channels, transporters, and other components of the cellular machinery controlling a broad range of cellular processes, including transcription, motility, contractility, and secretion. These cellular processes in turn regulate systemic functions such as embryonic development, gonadal development, learning and memory, and organismal homeostasis [2]. G protein-dependent and G protein-independent pathways each have the capacity to initiate numerous intracellular signaling cascades to mediate these effects [4]. G proteins are GTPases (guanosine

triphosphatases) that cycle between a GDP-bound form and a GTP-bound form [5]. The GTP-bound G protein is an active form that interacts with downstream effectors and transmits signals, during which the bound GTP is often hydrolyzed to GDP and the G protein recycles into the inactive GDP-bound form [5]. The heterotrimeric G protein complex comprises a Gα subunit, of which there are 4 main families (Gαs, Gαi/o, Gαq/11, and Gα12/13), coupled to a combination of Gβ and Gγ subunits, of which there exist 6 and 12 members, respecttively [2,4]. Gα subunit binds to guanine nucleotides while G$\beta\gamma$ subunits cannot be dissociated under nondenaturing conditions. The activity of G proteins is regulated mainly through three classes of regulatory proteins: GTPase-activating proteins (GAPs), guanine nucleotide-exchange factors (GEFs), and guanine nucleotide-dissociation inhibitors (GDIs) [6]. Upon activation, the GTP-bound Gα subunit dissociates from G$\beta\gamma$ subunits, and serves as the major signaling messenger by interacting with its signal acceptors (downstream effectors) [2].

Mammalian GPCRs constitute a superfamily of diverse proteins with hundreds of members [7,8]. GPCRs can be grouped into 6 classes based on sequence homology and functional similarity [9,10]: Class A (Rhodopsin-like receptors) [11], Class B (Secretin receptor family) [12], Class C (Metabotropic glutamate/pheromone receptors) [13], Class D (Fungal mating pheromone recaptors) [14], Class E (Cyclic AMP receptors) [15], and Class F (Frizzled/Smoothened, F/S) [16,17]. GPCRs act as receptors for a multitude of different signals [8]. One major group, referred to as chemosensory GPCRs (csGPCRs), is receptors for sensory signals of external origin that are sensed as odors [18,19], pheromones, or tastes [20]. Most other GPCRs respond to endogenous signals, such as peptides, lipids, neurotransmitters, or nucleotides [21,22]. These GPCRs are involved in numerous physiological processes, including the regulation of neuronal excitability, metabolism, reproduction, development, hormonal homeostasis, and behavior [8]. A characteristic feature of GPCRs differentially expressed in many cell types in the body, together with their structural diversity, has proved important in medicinal chemistry. GPCRs are involved in many diseases, and are also the target of around half of all modern medicinal drugs [23]. Of all currently marketed drugs, >30% are modulators of specific GPCRs [24]. However, only 10% of GPCRs are targeted by these drugs, emphasizing the potential of the remaining 90% of the GPCR superfamily for the treatment of human disease [8].

Additionally, Celera's initial analysis of the human genome found 616 GPCRs [25] and Takeda *et al.* [26] found 178 intronless nonchemosensory GPCRs, whereas the International Human Genome Sequencing Consortium reported a total of 569 "rhodopsin-like" (*i.e.*, Class A) GPCRs [27]. Vassilatis DK and co-worker conducted a comprehensive analysis and reported that the repertoire of GPCRs for endogenous ligands consists of 367 receptors in humans and 392 in mice. Included here are 26 human and 83 mouse GPCRs not previously identified [8]. Phylogenetic analyses cluster 60% of GPCRs according to ligand preference, allowing prediction of ligand types for dozens of orphan receptors. Expression profiling of 100 GPCRs demonstrates that most are expressed in multiple tissues and that individual tissues express multiple GPCRs. Over 90% of GPCRs are expressed in the brain. Strikingly, however, the profiles of most GPCRs are unique, yielding thousands of tissue- and cell-specific receptor combinations for the modulation of physiological processes.

Moreover, diverse members of GPCR superfamily participate in a variety of physiological functions and are major targets of pharmaceutical drugs. GPCRs are one of the most important target classes in pharmacology and are the target of many blockbuster drugs [28]. The presumably α-helical transmembrane regions (TMs) of GPCRs are probably arranged with similarity to bacteriorhodopsin (brh) [29]. Except for low-resolution electron diffraction [30,31] and high resolution X ray-based crystallography [32] of brh, the first crystal structure of a mammalian GPCR, bovine rhodopsin [33], was solved. In 2007, the first structure of a human GPCR, β_2-adrenergic receptor, was solved [34,35]. In particular, GPCRs are of enormous importance for the pharmaceutical industry because 52% of all existing medicines act on a GPCR [36]. Very well-known therapeutic drugs such as β-blockers and anti-histamines act on GPCRs. This explains why so many three-dimensional models of GPCRs have been built. Early structural models, such as HIV-1 co-receptor CCR5 (chemokine receptors) [37,38], and human thromboxane receptor [39], are based on the atomic coordinates of the brh structure; some models, e.g. human ADP receptor (Purinergic Receptor P2Y12) [40], are constructed by homology modeling using bovine rhodopsin as a template. All of these modeling studies combined with bioinformatics and chemoinformatics become amenable to the rational design of novel drugs targeting GPCRs in the human genome [28].

These models would contribute to a better understanding of the structure and the function of GPCRs, as well as the ligand-receptor interaction. The present study is devoted to use bioinformatics and computational modeling to build up GPCRs' theoretical modeling and folding fashions, for prediction of unknown GPCRs in the human genome and studying the interaction between GPCRs ant their ligands at the molecular level.

2. MATERIALS AND METHODS

Structural data of G-protein coupled receptors (GPCR)

were taken from a new release of the GPCRDB v.7.6 (http://www.gpcr.org/7tm/htmls/entries.html) based on the latest UniProtKB (Universal Protein Knowledgebase) release of 15-May-2006 (http://www.ebi.ac.uk/swissprot/; http://au.expasy.org/), which contain approximately 764 proteins. Their GPCR family profiles are updated. Their amino acid sequences were from Genbank (http://www.ncbi.nlm.nih.gov/Genbank/index.html) and SWISSPROT. The secondary structure of protein residues corresponds to the DSSP method and their seven TMs were determined based on the GPCR superfamily.

2.1. Data Partitioning

The transmembrane domain regions of 764 known GPCRs were each used as a query I TBLASTEN searches of the National Center for Biotechnology Information human genome database. Sequences were retrieved from the National Center for Biotechnology Information with the accession numbers (Appendix 1). GPCR Class A, B, and C Hidden Markov Model models were also used as queries to search the International Protein Index proteome database [8]. Grouping of the samples was based on the phylogenetic analysis results of Vassilatis and co-worker. Data sets were partitioned into three sets: Training, test, and validation sets. Although protein prediction methodology is almost always reported in terms of training and test sets only, we withheld an external validation set in order to provide an additional rigorous check on model quality. We feel this is necessary since a high statistical correlation on the training and test sets does not necessarily indicate a highly predictive model [41]. To properly partition our data sets so that they each reflect the makeup of the original data set as much as possible, we take into account the distribution of both feature diversity and biological activity as we form our training, test, and external validation sets. In this way, we maintain the original proportions of categorical bins and structural diversity in each of the three sets.

Training dataset is composed of 22 groups in three types of GPCRs (**Figure 1**) as follows: GPCRs from human different chromosomes (S_{DC}^{L}), from human same chromosomes (S_{SC}^{L}) (such as chromosome 3 and 11), and from different species, based on the phylogenetic trees [8]. The first contains five classes: Class A (S_{A}^{L}), B (S_{B}^{L}), C (S_{C}^{L}), F/S (S_{F}^{L}), and other (S_{O}^{L}). Class A consists of four groups: Group 1 (S_{A1}^{L}), Group 2 (S_{A2}^{L}), Group 3 (S_{A3}^{L}) and Group 4 (S_{A4}^{L}), which is abstracted into Group 14 ($S_{A14}^{L} \in S_{A}^{L}$). Class B, C, and F/S each contain one group (S_{B15}^{L}, S_{C16}^{L}, and S_{F22}^{L}). The others fall into nine groups $\left(S_{O5}^{L}, S_{O6}^{L}, \cdots, S_{O13}^{L}\right)$, which forms one group 17 $\left(S_{O17}^{L} \in S_{O}^{L}\right)$. The first is also extracted into one group 18 $\left(S_{D18}^{L} \in S_{DC}^{L}\right)$. The second consists of

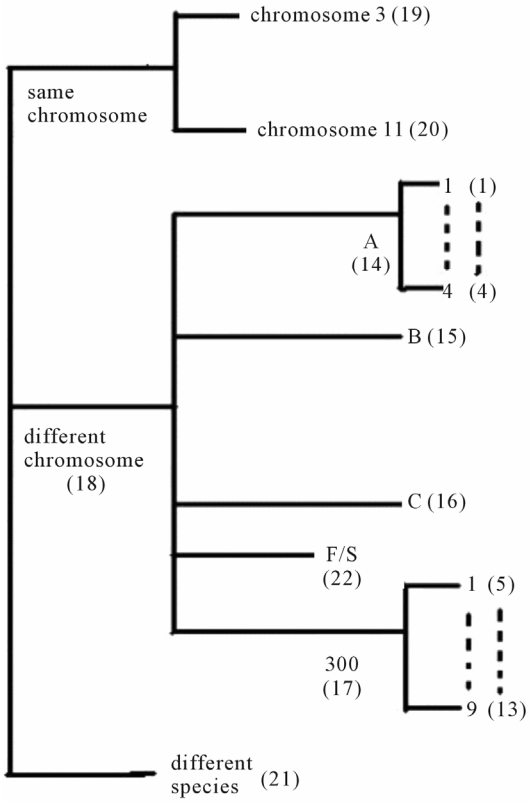

Figure 1. The grouping frame of the different training datasets.

two groups: chromosome 3 (group 19, S_{S19}^{L}) and chromosome 11 (group 20, S_{S20}^{L}). The third only includes one group 21 (S_{D21}^{L}), consisting of Bovine, Danre, Drome, Chick, Anoga, Dicla, Eisfo, Equas, Eulfu, Pantr, Halsh, besides human, consisting of 3, 2, 3, 3, 1, 1, 1, 1, 1, 1, 1, and 2 GPCRs, respectively. All above 22 groups make up of the training dataset (learning dataset, S^{L}).

The following test datasets $\left(S^{T}\right)$ contains two groups of GPCRs, group 23 from human (38 GPCRs, S_{23}^{T}) and group 24 $\left(S_{24}^{T}\right)$ from different species, consisting of 3 bovine, 3 canfa, 3 drome, 3 chick, 3 mouse, 1 Arath, 1 macmu, 1 Mesbi, and 1 Micoh GPCRs, respectively, besides 19 human GPCRs (Appendix 1). Here,

$$S_{A1}^{L} \cap S_{A2}^{L} \cap S_{A3}^{L} \cap S_{A4}^{L} = \Phi,$$

$$S_{A1}^{L} \cup S_{A2}^{L} \cup S_{A3}^{L} \cup S_{A4}^{L} = S_{A}^{L},$$

$$S_{A14}^{L} \in S_{A}^{L}, \quad S_{O5}^{L} \cap S_{O6}^{L} \cap \cdots \cap S_{O13}^{L} = \Phi,$$

$$S_{O17}^{L} \in S_{O}^{L}, \quad S_{O5}^{L} \cup S_{O6}^{L} \cup \cdots \cup S_{O13}^{L} = S_{O}^{L},$$

$$S_{A14}^{L} \cap S_{B15}^{L} \cap S_{C16}^{L} \cap S_{O17}^{L} \cap S_{F22}^{L} = \Phi,$$

and

$$S_{D18}^{L} \cap S_{S19}^{L} \cap S_{S20}^{L} \cap S_{D21}^{L} = \Phi.$$

The validation set involves 22 GPCRs from the sense chain of chromosome 19.

2.2. Sequence Analysis of GPCRs using Bioinformatics

2.2.1. The Scoring Matrices of Training Sets

Take Group 1 of Class A for an example. In order to represent the GPCRs' TM patterns, a representative non-redundant set of high resolution GPCRs' TMs are chosen as previously reported to build a training set (**Tables 1** and **2**). The most consistent sequences are picked up to constitute a scoring matrix by alignment that would be used to predict the TM regions. The amino acid sequences of the seven TMs of GPCRs were extracted and aligned using ClustalW; the TM regions cluster in one fragment (motif) which are about 12, 11, 13, 14, 10, 10, and 12 amino acid residues for TM1-TM7 of the Group 1 (**Table 2**), respectively; and then their coding regions of such amino acid fragments were chosen to constitute the scoring matrix, which contains 4 types of nucleotides (**Figure 2**).

Take TM1 of GPCRs in Group 1 of Class A for an example. There are 42 GPCR proteins consisting of the training set after alignment (**Table 1**). **Figure 2** means the scoring matrix, which was generated by assigning a value of the stimulatory potential to each of the 4 defined nucleotides in each position of **Table 1**. Based on the matrix, we designed a simple algorithm to evaluate the relationship significance of any sequence to the GPCRs' TM patterns. To each nucleotide n (A, T, G, and C) from those 42 proteins (i) of TM1 of group 1 (**Table 2**), the symbol tij stands for how many times it takes place in each position $j\left(j \in 1,2,3,\cdots,12\times 3\right)$, which was calculated as follows: $tij = \sum_{i=1}^{42} nij$. The score Sij of this nucleotide denotes the proportional (weighting) it takes place in each position j, which was calculated as follows: $Sij = tij \Big/ \left(\sum_{n=1}^{4} \sum_{j=1}^{12\times 3} \sum_{i=1}^{42} nij \right)$. Take the adenosine (A) for example. Based on the **Table 1**, the times of adenosine is $5,12,\cdots$, at the position of $1,2,\cdots$, respectively, and the sum of four nucleotides in the training set is 1512 $\left(36\times 42 = 1512\right)$. So, the scores $\left(Sij\left(\mathrm{A}\right)\right)$ of adenosine is $0.003, 0.008,\cdots$ at the position of $1,2,\cdots$, respectively, whereas it is 0 at other position because it does not appear (**Figure 2**). The rest (Thymine, Cytidine, and Guanosine) may be deduced by analogy. The value of the scoring matrix is 1.

2.2.2. Test Sets

According to the set theory of mathematics [42], the GPCRs chosen above consist of different training sets S_A^L, S_B^L, S_C^L, S_F^L, S_O^L, etc, which composed a union $S_A^L \bigcup S_B^L \bigcup S_C^L \bigcup \cdots = S^L$, and $S_A^L \bigcap S_B^L \bigcap S_C^L \bigcap \cdots = \Phi$. Therefore, the test set $\left(S^T\right)$ (**Table 3**) comes from the complement of S^L for GPCRs aggregate (Appendix 1).

According to our previous methods [40,43], we defined the coding sequence (CDS) of GPCRs' each TM as TM-CDS unit composed of m nucleotides. At first, the TM-CDS units are obtained using the sliding window method one by one from 5'-terminal of GPCRs' CDS to 3'-teminal: A sequence of l nucleotides gives rise to $l-m+1$ TM-CDS units. For example, the coding sequences of TM1 of GPCRs in group 1 are 12 × 3 nucleotides, namely $m = 36$.

2.2.3. Validation Set

Similarly, we calculate the total scores of the coding sequences of 22 GPCRs located at the sense chain of chromosome 19 using the sliding window method.

2.2.4. Assessment of Model Quality

In this study, training model quality is simply the percent correct classification (binning) of GPCRs' TM segments for the test set [41]. The overall predictive power of a given model is the percent correct classification for the test set (%test) and for the external validation set (%validation), where the external validation set represents native holdout data. More extensive model assessment was accomplished by a "dynamic partitioning" procedure, which provides a no error rate of the test and external validation sets.

2.2.5. Statistics

Data are expressed as mean±standard deviation (S.D.) through this paper. Statistical analyses were performed with *F*-test by one-way analysis of variance (abbreviated one-way ANOVA) and by t-test between the means of two groups of the samples. Data was considered significant for $P < 0.01$ at 95 confidence limit [44]. Tests for normality were performed with Shapiro-Wilk test because of the number of samples less than 2000 [45]. The normality of the data was tested by the Shapiro-Wilk statistic. All statistical testing was conducted at significance level 0.10 and all confidence intervals had confidence level 0.90 unless otherwise noted. All tests and confidence intervals were two-sided. Confidence intervals for normal data were constructed from analysis of covariance models [45]. Here, α = 0.10 requests 90% confidence limits. The default value is 0.05. One way-ANOVA, Test of Homogeneity of Variances and Multiple comparisons (LSD and Tamhane's T2), and tests for normality were performed using SPSS version 11.5 software.

2.2.6. The Prediction Model Algorithm

In general, our prediction model (PreMod) method employs the scoring matrices combined with descriptor

Table 1. TM1 sequence alignment of GPCRs in group 1 of class A by clustal W.

GPCRs	TM1			
	Interzone	Amino acid sequences	Alignment motif	The coding sequences
OX1R_HUMAN	47 - 67	WVLIAAYVAVFVVALVGNTLV	YVAVFVVALVGN	tatgtggctgtgttcgtcgtggccctggtgggcaac
OX2R_HUMAN	55 - 75	WVLIAGYIIVFVVALIGNVLV	YIIVFVVALIGN	tacatcatcgtgttcgtcgtggctctcattgggaac
NPFF2_HUMAN	148 - 168	AIFIISYFLIFFLCMMGNTVV	YFLIFFLCMMGN	tactttctgatcttcttttgtgcatgatgggaaat
NPFF1_HUMAN	44 - 64	AMFIVAYALIFLLCMVGNTLV	YALIFLLCMVGN	tatgcgctcatcttcctgctctgcatggtgggcaac
GPR83_HUMAN	72 - 92	ALLIVAYSFIIVFSLFGNVLV	YSFIIVFSLFGN	tactccttcatcattgtcttctcactctttggcaac
NPY1R_HUMAN	45 - 65	LAYGAVIILGVSGNLALIIII	YGAVIILGVSGN	tatggagctgtgatcattcttggtgtctctggaaac
NPY2R_HUMAN	52 - 72	VLILAYCSIILLGVIGNSLVI	YCSIILLGVIGN	tactgctccatcatcttgcttggggtaattggcaac
NPY4R_HUMAN	40 - 60	VMVFIVTSYSIETVVGVLGNL	YSIETVVGVLGN	tacagcattgagactgtcgtgggggtcctgggtaac
CCKAR_HUMAN	42 - 67	AVQILLYSLIFLLSVLGNTLVITVLI	YSLIFLLSVLGN	tactccttgatattcctgctcagcgtgctgggaaac
GASR_HUMAN	58 - 79	ITLYAVIFLMSVGGNMLIIVVL	YAVIFLMSVGGN	tacgcagtgatcttcctgatgagcgttggaggaaat
NMBR_HUMAN	42 - 65	IRCVIPSLYLLIITVGLLGNIMLV	YLLIITVGLLGN	tacctgctcatcatcaccgtgggcttgctgggcaac
GRPR_HUMAN	39 - 62	ILYVIPAVYGVIILIGLIGNITLI	YGVIILIGLIGN	tatggggttatcattctgataggcctcattggcaac
EDNRA_HUMAN	81 - 102	YINTVISCTIFIVGMVGNATLL	SCTIFIVGMVGN	tcttgtactattttcatcgtgggaatggtggggaat
EDNRB_HUMAN	102 - 126	YINTVVSCLVFVLGIIGNSTLLRII	SCLVFVLGIIGN	tcctgccttgtgttcgtgctggggatcatcgggaac
LGR6_HUMAN	568 - 588	VWAIVLLSVLCNGLVLLTVFA	VWAIVLLSVLCN	gtgtgggccatcgtgttgctctccgtgctctgcaat
LGR4_HUMAN	545 - 565	VWFIFLVALFFNLLVILTTFA	VWFIFLVALFFN	gtgtggttcattttcttggttgcattatttttcaac
LGR5_HUMAN	562 - 582	IGVWTIAVLALTCNALVTSTV	VWTIAVLALTCN	gtgtggaccatagcagttctggcacttacttgtaat
V1BR_HUMAN	36 - 59	VEIGVLATVLVLATGGNLAVLLTL	LATVLVLATGGN	ctggccactgtcctggtgctggcgaccgggggcaac
V1AR_HUMAN	53 - 76	LEIAVLAVTFAVAVLGNSSVLLAL	LAVTFAVAVLGN	ctggcggtgactttcgcggtggccgtgctgggcaac
GNRHR_HUMAN	39 - 58	VTVTFFLFLLSATFNASFLL	FLFLLSATFNAS	ttccttttctgctctctgcgacctttaatgcttct
GPR27_HUMAN	24 - 44	TLSLLLCVSLAGNVLFALLIV	LSLLLCVSLAGN	ctcagcctgctgctgtgcgtgagcctagcgggcaac
GPR85_HUMAN	26 - 46	SLGFIIGVSVVGNLLISILLV	LGFIIGVSVVGN	ttgggtttcataataggagtcagcgtggtgggcaac
OPN3_HUMAN	41 - 65	YERLALLLGSIGLLGVGNNLLVLVL	LGSIGLLGVGNN	ctgggctccattgggctgctgggcgtcggcaacaac
OPSX_HUMAN	27 - 49	IVATYLIMAGMISIISNIIVLGI	LIMAGMISIISN	ttgattatggcaggtatgataagtattatcagcaac
RGR_HUMAN	16 - 36	VLAVGMVLLVEALSGLSLNTL	VGMVLLVEALSG	gtggggatggtgctactggtggaagctctctccggt
OPSD_HUMAN	37 - 61	FSMLAAYMFLLIVLGFPINFLTLYV	MFLLIVLGFPIN	atgtttctgctgatcgtgctgggcttccccatcaac
OPSG_HUMAN	53 - 77	YHLTSVWMIFVVIASVFTNGLVLAA	MIFVVIASVFTN	atgatctttgtggtcattgcatccgttttcacaaat
OPSR_HUMAN	53 - 77	YHLTSVWMIFVVTASVFTNGLVLAA	MIFVVTASVFTN	atgatctttgtggtcactgcatccgtcttcacaaat
MTR1A_HUMAN	30 - 50	LACVLIFTIVVDILGNLLVIL	LIFTIVVDILGN	ctcatcttcaccatcgtggtggacatcctgggcaac
MTR1B_HUMAN	43 - 63	LSAVLIVTTAVDVVGNLLVIL	LIVTTAVDVVGN	ctcatcgtcaccaccgccgtggacgtcgtgggcaac
EDG1_HUMAN	47 - 71	LTSVVFILICCFIILENIFVLLTIW	FILICCFIILEN	ttcattctcatctgctgctttatcatcctggagaac
EDG3_HUMAN	41 - 65	LTTVLFLVICSFIVLENLMVLIAIW	FLVICSFIVLEN	ttcttggtcatctgcagcttcatcgtcttggagaac
EDG5_HUMAN	35 - 59	VASAFIVILCCAIVVENLLVLIAVA	IVILCCAIVVEN	atcgtcatcctctgttgcgccattgtggtggaaaac
EDG6_HUMAN	51 - 71	GLSVAASCLVVLENLLVLAAI	SVAASCLVVLEN	tcggtggccgccagctgcctggtggtgctggagaac
CNR2_HUMAN	34 - 59	TAVAVLCTLLGLLSALENVAVLYLIL	CTLLGLLSALEN	tgcactcttctgggcctgctaagtgccctggagaac
CNR1_HUMAN	117 - 142	LAIAVLSLTLGTFTVLENLLVLCVIL	SLTLGTFTVLEN	tccctcacgctgggcaccttcacggtcctggagaac
MC5R_HUMAN	38 - 61	IAVEVFLTLGVISLLENILVIGAI	FLTLGVISLLEN	tttctcactctgggtgtcatcagcctcttggagaac
MC4R_HUMAN	44 - 69	LFVSPEVFVTLGVISLLENILVIVAI	FVTLGVISLLEN	tttgtgactctgggtgtcatcagcttgttggagaat
PE2R1_HUMAN	36 - 62	PALPIFSMTLGAVSNLLALALLAQAAG	PIFSMTLGAVSN	cccatcttctccatgacgctgggcgccgtgtccaac
PE2R4_HUMAN	20 - 43	PVTIPAVMFIFGVVGNLVAIVVLC	PAVMFIFGVVGN	ccggcggtgatgttcatcttcggggtggtgggcaac
PE2R2_HUMAN	24 - 47	SPAISSVMFSAGVLGNLIALALLA	SSVMFSAGVLGN	agctccgtcatgttctcggccggggtgctggggaac
PD2R_HUMAN	22 - 42	MGGVLFSTGLLGNLLALGLLA	GGVLFSTGLLGN	ggcggggtgctcttcagcaccggcctcctgggcaac

Table 2. The amino acid sequence length and the sample number of the scoring matrix in the training datasets after sequence alignments.

Class	Training Datasets		TM1		TM2		TM3		TM4		TM5		TM6		TM7	
	S^L	Number	Re*	Sp*	Re*	Sp*	Re*	Sp*	Re*	Sp*	Re*	Sp*	Re*	Sp*	Re*	Sp*
A	S^L_{A1}	44	12	42	11	44	13	43	14	43	10	41	10	41	12	43
	S^L_{A2}	38	16	37	13	38	15	36	12	35	10	35	10	38	14	34
	S^L_{A3}	32	17	30	15	32	14	32	10	30	14	31	16	32	13	29
	S^L_{A4}	20	15	19	18	18	14	20	12	20	11	19	12	20	13	20
	S^L_{A14}	44	13	43	11	35	13	44	11	42	13	35	13	42	11	41
B	S^L_{B15}	13	12	13	12	13	11	13	9	13	16	13	13	13	13	13
C	S^L_{C16}	10	15	10	10	10	16	10	20	10	12	10	20	10	15	10
F/S	S^L_{F22}	9	16	9	21	9	19	9	20	9	19	9	21	9	18	9
O	S^L_{O5}	39	21	39	21	38	18	38	21	39	10	39	12	39	21	39
	S^L_{O6}	21	21	24	19	24	18	24	15	24	21	24	16	24	20	24
	S^L_{O7}	33	20	33	17	33	20	33	20	33	10	32	20	33	12	33
	S^L_{O8}	11	21	11	18	11	20	10	19	11	20	11	21	11	20	11
	S^L_{O9}	48	14	48	18	48	9	48	13	48	15	48	13	48	20	48
	S^L_{O10}	40	24	40	22	40	20	39	19	40	24	40	23	40	20	40
	S^L_{O11}	44	14	44	13	44	18	44	11	44	12	43	19	43	20	44
	S^L_{O12}	20	19	20	18	20	19	20	15	20	20	20	19	20	21	20
	S^L_{O13}	20	18	20	15	20	21	20	15	20	18	20	21	20	21	20
	S^L_{O17}	33	13	33	18	33	16	33	15	28	13	31	20	31	20	33
Chromosome	S^L_{D18}	39	16	33	13	35	14	35	11	36	12	34	12	35	13	29
	S^L_{S19}	27	17	22	14	24	12	25	15	27	14	26	15	26	12	25
	S^L_{S20}	22	13	17	11	21	17	22	14	22	15	15	10	22	11	20
Species	S^L_{D21}	20	12	19	14	13	18	20	12	20	12	18	13	20	9	17

Note: "Number" means the total sample numbers of each training dataset; "Re", the amino acid residue length of each transmembrane region; "Sp", the actual sample numbers of each transmembrane region in each training dataset.

selection procedures (seven TMs) that seek to find the optimal subset of the scoring matrices from the original scoring matrix manifold. Partitioning data sets into training, test, and external validation sets rigorously assesses model quality. We extend this methodology by implementation of the dynamic repeating assessment. A flowchart of the prediction model algorithm is provided in **Figure 3**, which involves the following steps. 1) Divide each data set into two parts: One used to build models, the other to validate models (external validation set); in our implementation, the external validation set is selected to have a high level of diversity; 2) Further partition the 80% identified for model building to form two more sets: Training (80%) and test (20%) sets; 3) Select seven TMs of GPCRs as descriptors based on phylogenetic evolution of the training set with or without cross-validation procedure (described above); 4) Calculate the score *Sij* of the training set to construct an optimized subset of the scoring matrix based on the CDS of GPCRs' 7 TMs; 5) Predict the test set target values using the scoring matrix and calculate the percent correct classification of the test set (%test); 6) Merge the training and test sets, and build a new prediction model using statistic analyses; 7) Predict external validation set values using the prediction model (PreMod), and calculate the percent correct classification of the external validation set (%validation); 8) Repeat steps 1-8 a preset number of times (22 times); 9) Assess each model by the accuracy described above, and generate test and external validation veracity.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	0.003	0.008	0	0.007	0	0.001	0.007	0	0	0.014	0.001	0.003	0.008	0	0.002	0.007	0	0.001	0.004	0	0.003
T	0.016	0.014	0.005	0.008	0.013	0.005	0.007	0.02	0.009	0.001	0.023	0.003	0.011	0.019	0.005	0.007	0.018	0.004	0.004	0.024	0.003
C	0.005	0.004	0.014	0.003	0.006	0.012	0.005	0.008	0.012	0.006	0.004	0.011	0.003	0.002	0.018	0.005	0.005	0.011	0.009	0.004	0.009
G	0.003	0.002	0.009	0.01	0.009	0.009	0.008	0	0.007	0.007	0	0.011	0.007	0.007	0.003	0.009	0.005	0.012	0.011	0	0.013

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
A	0.009	0.002	0.003	0.005	0	0.002	0.005	0.001	0.001	0.003	0.006	0.005	0.026	0.026	0
T	0.004	0.003	0.003	0.003	0.025	0.004	0.005	0.022	0.005	0.003	0.001	0.002	0.001	0	0.007
C	0	0.008	0.016	0.005	0.003	0.013	0.009	0.003	0.005	0	0.003	0.014	0	0.001	0.021
G	0.015	0.015	0.005	0.014	0	0.009	0.008	0.002	0.017	0.021	0.017	0.007	0.001	0.001	0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	0.003	0	0.003	0.003	0	0.005	0.001	0	0.002	0.008	0.001	0.007	0.003	0.03	0	0.006	0	0.004	0.003	0	0.003
T	0.005	0.03	0.004	0.005	0.002	0.009	0.006	0.024	0.004	0.005	0.001	0.01	0	0	0.006	0.005	0.028	0.003	0.006	0.024	0.001
C	0.022	0	0.004	0.001	0.025	0.01	0.008	0.003	0.012	0.001	0.021	0.01	0.001	0.001	0.024	0.018	0.001	0.006	0.015	0.004	0.008
G	0	0	0.019	0.021	0.003	0.006	0.016	0.003	0.012	0.017	0.008	0.003	0.027	0	0.001	0.002	0.002	0.018	0.007	0.002	0.019

	22	23	24	25	26	27	28	29	30	31	32	33
A	0.007	0.003	0.002	0.012	0.002	0.005	0.009	0.001	0.004	0.011	0.003	0.001
T	0.001	0.017	0.003	0.003	0.004	0.005	0.004	0.025	0.004	0.01	0.017	0.009
C	0.004	0.006	0.009	0.002	0.012	0.015	0.006	0.003	0.015	0.003	0.01	0.017
G	0.019	0.005	0.016	0.014	0.012	0.004	0.011	0.002	0.007	0.006	0.001	0.003

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	0.007	0.002	0.002	0.007	0	0.001	0.003	0	0.002	0.004	0	0.003	0.007	0.005	0.006	0.004	0.001	0.007	0.004	0	0.003
T	0.008	0.011	0.008	0.008	0.02	0.004	0.018	0.005	0.007	0.005	0.016	0.002	0.007	0.01	0.007	0.002	0.009	0.004	0.017	0	0.007
C	0.002	0.006	0.013	0.007	0.004	0.012	0	0.016	0.014	0.003	0.002	0.007	0.004	0.007	0.005	0.001	0.013	0.011	0.001	0.023	0.013
G	0.009	0.007	0.004	0.005	0.001	0.01	0.005	0.005	0.003	0.014	0.007	0.013	0.008	0.005	0.008	0.018	0.003	0.004	0.004	0.003	0.003

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
A	0.01	0	0.003	0.002	0.001	0.002	0.018	0.003	0.004	0.003	0	0.004	0.009	0	0.003	0.001	0	0.002
T	0	0.018	0.005	0.015	0.018	0.002	0.004	0.008	0.005	0.004	0.025	0.001	0.004	0.014	0.006	0.004	0.011	0.003
C	0.004	0.007	0.01	0.007	0.001	0.016	0.004	0.009	0.009	0.018	0	0.011	0.006	0.008	0.008	0.004	0.012	0.017
G	0.013	0	0.007	0.002	0.006	0.005	0	0.006	0.007	0.001	0.001	0.01	0.007	0.004	0.008	0.017	0.003	0.004

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	0.014	0	0.001	0.004	0	0.002	0.003	0.001	0.001	0.013	0	0.001	0.001	0.001	0	0.006	0	0.003	0.006	0.002	0.002
T	0.001	0.022	0.008	0.003	0.009	0.007	0.006	0.013	0.003	0.004	0.015	0.003	0.022	0.002	0.002	0.001	0.015	0.003	0.004	0.018	0.006
C	0.004	0.002	0.011	0.004	0.009	0.01	0.004	0.006	0.015	0.001	0.006	0.016	0.001	0.001	0.002	0.006	0.008	0.012	0.009	0.002	0.007
G	0.005	0	0.004	0.012	0.006	0.006	0.011	0.004	0.004	0.006	0.003	0.004	0.001	0.02	0.02	0.012	0.001	0.006	0.005	0.002	0.009

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
A	0.002	0.001	0.002	0.004	0.001	0.002	0.005	0	0.003	0.004	0.001	0.002	0.007	0	0.004	0.005	0.001	0.002	0.002	0	0.003
T	0.011	0	0.006	0.006	0.015	0.004	0.003	0.016	0.003	0.009	0.015	0.002	0.004	0.005	0.003	0.006	0.013	0.007	0.001	0.011	0.004
C	0.001	0.02	0.013	0.007	0.006	0.004	0.004	0.007	0.011	0.006	0.004	0.009	0	0.012	0.012	0.003	0.008	0.007	0.019	0.013	0.006
G	0.011	0.003	0.003	0.008	0.002	0.014	0.011	0.001	0.007	0.004	0.004	0.011	0.013	0.007	0.005	0.009	0.003	0.008	0.002	0.001	0.011

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	0.002	0.008	0.001	0.01	0.001	0.001	0.013	0.002	0.002	0.004	0	0.011	0.003	0	0.002	0.007	0	0.002	0.013	0.003	0.001
T	0.024	0.022	0.004	0.011	0.023	0.005	0.002	0.02	0.005	0.002	0.008	0.004	0.004	0.028	0.006	0.01	0.013	0.009	0.007	0.024	0.004
C	0.005	0.001	0.025	0.007	0.007	0.018	0.005	0.009	0.017	0.023	0.024	0.014	0.022	0.003	0.011	0.003	0.011	0.015	0.007	0.005	0.024
G	0.002	0.002	0.003	0.006	0.003	0.01	0.013	0.002	0.009	0.004	0.002	0.005	0.004	0.002	0.015	0.013	0.009	0.007	0.007	0.001	0.004

	22	23	24	25	26	27	28	29	30
A	0.024	0	0.004	0.007	0	0.002	0.009	0.004	0
T	0.002	0.03	0.004	0.007	0.024	0.007	0.011	0.025	0.01
C	0.005	0.003	0.014	0.003	0.008	0.01	0.002	0.003	0.015
G	0.003	0	0.011	0.016	0.002	0.014	0.011	0.001	0.008

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	0.002	0.003	0	0.008	0	0.004	0.009	0.002	0.001	0.001	0	0.002	0	0.002	0	0.005	0	0.004	0	0	0.008
T	0.028	0.029	0.014	0.005	0.019	0.004	0.005	0.028	0.004	0.029	0.002	0.004	0.031	0	0.002	0.007	0.02	0.005	0	0	0.01
C	0.002	0.001	0.017	0.009	0.012	0.019	0.007	0.003	0.023	0.002	0.003	0.027	0.002	0.004	0.003	0.008	0.007	0.011	0.031	0.033	0.011
G	0.002	0	0.002	0.011	0.002	0.007	0.012	0	0.006	0.002	0.028	0.001	0	0.028	0.028	0.014	0.006	0.013	0.002	0	0.004

	22	23	24	25	26	27	28	29	30
A	0.006	0.008	0	0.007	0.008	0.001	0.012	0	0.003
T	0.013	0.024	0.01	0.013	0.016	0.007	0.007	0.02	0.004
C	0.012	0.001	0.019	0.009	0.006	0.02	0.002	0.006	0.019
G	0.002	0	0.005	0.004	0.003	0.007	0.012	0.007	0.007

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	0.004	0.001	0.002	0.003	0	0.006	0.01	0.01	0.007	0.008	0	0.002	0.017	0.017	0.001	0.01	0.001	0.005	0.01	0	0.001
T	0.009	0.026	0.01	0.004	0.006	0.006	0.012	0.011	0.005	0.008	0.015	0.005	0.004	0.001	0.009	0.01	0	0.006	0.012	0.01	0.006
C	0.013	0.001	0.003	0.001	0.017	0.014	0.003	0.007	0.01	0.006	0.006	0.012	0.003	0.009	0.016	0.003	0.021	0.016	0.001	0.005	0.016
G	0.002	0	0.012	0.019	0.004	0.002	0.003	0	0.006	0.006	0.006	0.008	0.003	0.001	0.002	0.005	0.006	0.002	0.005	0.014	0.004

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
A	0.006	0.005	0.002	0.023	0.027	0	0	0	0.005	0.01	0	0.001	0.018	0.001	0.004
T	0.006	0.019	0.004	0	0	0.008	0.001	0	0.005	0.008	0.023	0.007	0.001	0.025	0.008
C	0.005	0.003	0.016	0	0.001	0.019	0.025	0.028	0.017	0.005	0	0.012	0.003	0.003	0.014
G	0.01	0.001	0.006	0.005	0	0	0.002	0	0.001	0.006	0.005	0.007	0.006	0	0.003

Figure 2. The scoring matrices of seven transmembrane regions of GPCRs in Group 1 of the training datasets (TM1-TM7: From top to bottom).

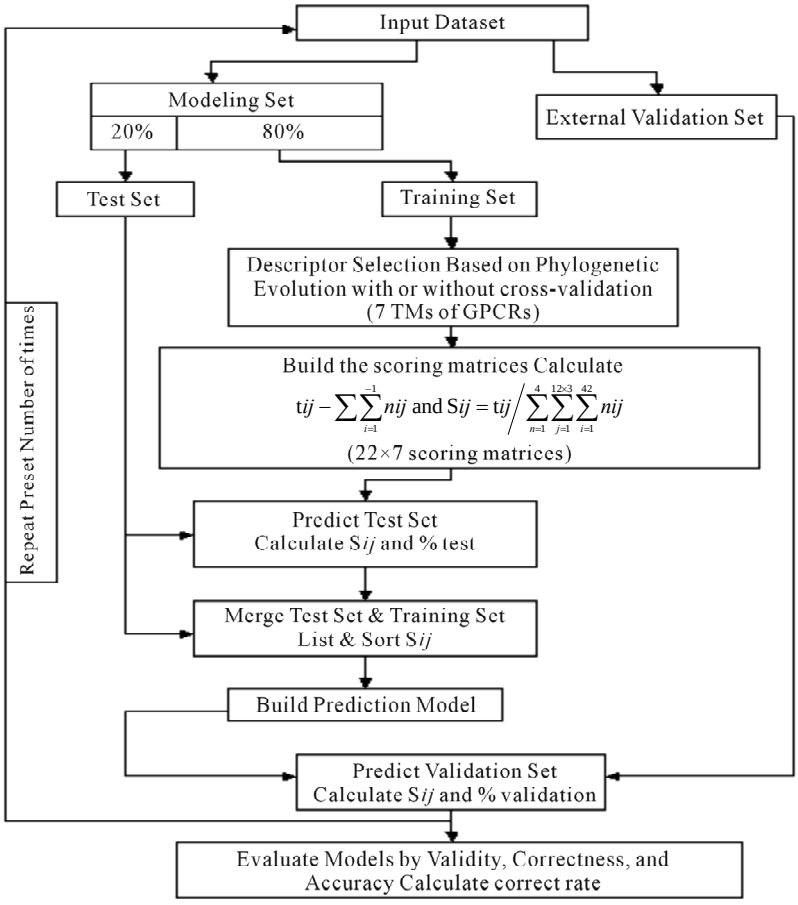

Figure 3. Flowchart of the algorithrm.

Table 3. The scores and the validity of prediction each transmembrane region of GPCRs in test sets by the scoring matrices of training set 18, 14 and 3.

GPCRs	S^L_{D18}														S^L_{A14}														S^L_{A3}													
	Score							Validity*							Score							Validity*							Score							Validity*						
	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7
Test set 23																																										
Q5VTM0_HUMAN	0.473	0.494	0.386	0.424	0.424	0.437	0.458	1	2	0	2	2	2	2	0.527	0.546	0.427	0.423	0.435	0.457	0.394	1	2	2	2	0	2	2	0.531	0.495	0.462	0.429	0.456	0.435	0.475	1	2	0	0	0	2	2
NPY5R_HUMAN	0.424	0.464	0.362	0.387	0.384	0.426	0.434	2	2	0	2	0	2	2	0.436	0.538	0.371	0.395	0.385	0.467	0.371	2	2	0	2	0	2	0	0.478	0.483	0.398	0.422	0.415	0.443	0.531	1	2	0	2	0	2	2
BRS3_HUMAN	0.398	0.451	0.350	0.391	0.441	0.498	0.415	1	2	0	2	2	2	2	0.465	0.505	0.372	0.462	0.444	0.553	0.419	1	2	0	2	2	2	2	0.475	0.467	0.402	0.489	0.435	0.549	0.490	1	2	0	2	2	2	2
LGR8_HUMAN	0.392	0.418	0.390	0.422	0.405	0.425	0.454	2	0	2	0	0	2	1	0.418	0.462	0.409	0.450	0.419	0.468	0.395	2	0	2	0	2	2	0	0.426	0.423	0.411	0.479	0.437	0.420	0.473	1	0	0	0	2	2	1
GP173_HUMAN	0.391	0.406	0.378	0.406	0.411	0.454	0.454	2	2	2	2	0	2	2	0.456	0.474	0.402	0.455	0.423	0.487	0.431	2	0	0	2	0	2	0	0.427	0.444	0.422	0.470	0.427	0.495	0.506	2	0	1	2	0	2	2
OPSB_HUMAN	0.409	0.461	0.395	0.382	0.484	0.439	0.453	2	2	0	2	2	2	2	0.457	0.510	0.434	0.429	0.500	0.466	0.433	2	2	0	0	2	2	0	0.473	0.472	0.475	0.443	0.524	0.508	0.487	2	2	0	2	2	2	2
EDG8_HUMAN	0.425	0.444	0.399	0.407	0.419	0.468	0.514	0	2	0	0	2	2	2	0.472	0.499	0.451	0.460	0.433	0.491	0.425	0	2	0	2	2	2	2	0.474	0.461	0.430	0.499	0.443	0.467	0.562	0	0	0	2	2	2	2
PF2R_HUMAN	0.386	0.373	0.356	0.363	0.389	0.414	0.383	2	0	0	2	0	0	2	0.442	0.435	0.367	0.393	0.383	0.470	0.369	2	0	0	0	0	1	0	0.437	0.394	0.432	0.458	0.396	0.462	0.409	2	0	0	2	0	2	0
ADA2C_HUMAN	0.458	0.464	0.395	0.437	0.432	0.547	0.539	2	2	0	2	2	2	2	0.555	0.552	0.435	0.486	0.487	0.602	0.469	2	2	0	2	2	2	2	0.514	0.450	0.442	0.472	0.455	0.554	0.602	2	2	0	2	2	2	2
ADRB3_HUMAN	0.407	0.442	0.405	0.401	0.473	0.518	0.505	2	2	2	2	2	2	2	0.479	0.509	0.440	0.457	0.484	0.562	0.431	2	2	2	2	2	2	2	0.467	0.443	0.417	0.469	0.459	0.507	0.526	2	0	2	0	2	2	2
5HT1E_HUMAN	0.398	0.466	0.386	0.426	0.428	0.480	0.502	0	2	2	2	0	2	2	0.423	0.556	0.470	0.469	0.457	0.529	0.400	0	2	2	2	0	2	0	0.423	0.449	0.445	0.511	0.490	0.473	0.501	0	2	2	2	0	2	2
5HT6R_HUMAN	0.398	0.420	0.434	0.396	0.424	0.478	0.536	0	2	2	2	2	2	1	0.455	0.486	0.492	0.439	0.458	0.530	0.452	2	2	2	2	2	2	0	0.456	0.435	0.461	0.489	0.451	0.493	0.582	0	0	2	2	2	2	1
TAAR2_HUMAN	0.363	0.461	0.408	0.372	0.404	0.440	0.446	0	2	0	0	0	2	2	0.404	0.551	0.403	0.399	0.428	0.492	0.417	0	2	0	0	2	2	0	0.431	0.500	0.419	0.420	0.428	0.456	0.463	0	2	0	0	0	2	2
AA3R_HUMAN	0.400	0.447	0.406	0.376	0.421	0.434	0.508	2	2	2	2	0	2	2	0.504	0.510	0.446	0.416	0.455	0.473	0.451	2	2	2	2	2	2	0	0.473	0.444	0.449	0.474	0.478	0.436	0.542	2	0	0	2	2	2	2
ACM5_HUMAN	0.446	0.449	0.403	0.429	0.432	0.460	0.498	2	2	0	2	2	2	2	0.497	0.491	0.438	0.434	0.498	0.504	0.441	2	2	2	2	2	2	0	0.478	0.452	0.456	0.474	0.465	0.504	0.530	2	2	0	0	2	2	2
HRH4_HUMAN	0.420	0.468	0.385	0.449	0.402	0.450	0.471	0	2	0	2	2	2	2	0.425	0.539	0.425	0.452	0.419	0.478	0.407	0	2	0	2	0	2	0	0.429	0.506	0.434	0.487	0.435	0.460	0.479	0	2	0	2	2	2	2
CXCR6_HUMAN	0.436	0.490	0.420	0.428	0.468	0.443	0.473	2	2	2	2	2	2	2	0.496	0.575	0.477	0.491	0.467	0.506	0.388	2	2	2	2	2	2	1	0.571	0.567	0.522	0.580	0.518	0.541	0.555	2	2	2	1	2	2	2
SSR5_HUMAN	0.422	0.495	0.402	0.416	0.463	0.484	0.512	2	2	2	2	2	2	2	0.475	0.602	0.459	0.468	0.486	0.530	0.454	2	2	2	2	2	2	2	0.542	0.591	0.570	0.486	0.537	0.536	0.636	2	2	2	2	2	2	2
OPRX_HUMAN	0.448	0.490	0.433	0.429	0.496	0.484	0.571	2	2	2	2	2	2	2	0.484	0.594	0.475	0.465	0.505	0.549	0.490	2	2	2	2	2	2	2	0.566	0.583	0.544	0.471	0.520	0.565	0.685	2	2	2	2	2	2	2
FPR1_HUMAN	0.457	0.470	0.392	0.418	0.428	0.456	0.560	2	2	2	2	2	2	2	0.554	0.552	0.439	0.447	0.459	0.497	0.475	2	2	2	2	2	2	2	0.572	0.558	0.545	0.541	0.482	0.515	0.697	2	2	2	2	2	2	2
C5ARL_HUMAN	0.475	0.476	0.450	0.412	0.423	0.516	0.486	2	2	1	2	2	2	2	0.546	0.581	0.473	0.452	0.461	0.568	0.449	2	2	2	2	2	2	2	0.559	0.562	0.507	0.541	0.525	0.566	0.640	2	2	2	2	2	1	2
GALR3_HUMAN	0.461	0.502	0.429	0.404	0.400	0.466	0.538	2	2	2	2	2	0	1	0.549	0.604	0.454	0.453	0.455	0.518	0.521	2	2	2	2	2	0	1	0.561	0.583	0.490	0.478	0.488	0.525	0.656	2	2	2	2	2	2	1
P2Y12_HUMAN	0.422	0.415	0.383	0.396	0.413	0.441	0.461	1	0	0	2	0	2	2	0.441	0.455	0.388	0.429	0.425	0.486	0.407	1	0	0	2	0	1	2	0.521	0.462	0.456	0.454	0.416	0.504	0.473	1	2	2	2	0	2	2
P2RY4_HUMAN	0.410	0.442	0.400	0.380	0.397	0.494	0.508	0	2	0	0	2	2	2	0.423	0.510	0.443	0.422	0.443	0.546	0.445	0	2	2	0	0	2	2	0.438	0.531	0.536	0.428	0.455	0.522	0.606	2	2	2	0	0	2	2
PAR4_HUMAN	0.439	0.481	0.387	0.391	0.437	0.421	0.529	2	2	0	0	2	2	2	0.509	0.565	0.466	0.427	0.467	0.439	0.457	2	2	2	0	2	2	2	0.548	0.542	0.443	0.466	0.507	0.447	0.573	2	1	2	0	2	2	2
MRGRD_HUMAN	0.410	0.509	0.425	0.407	0.395	0.456	0.462	1	2	0	0	0	2	2	0.483	0.604	0.410	0.426	0.407	0.518	0.411	2	2	0	0	0	1	0	0.491	0.599	0.461	0.469	0.449	0.466	0.539	1	2	0	2	2	2	2
EMR1_HUMAN	0.392	0.401	0.402	0.402	0.417	0.430	0.444	0	0	0	0	0	0	1	0.446	0.473	0.414	0.421	0.419	0.439	0.418	0	0	0	0	0	0	0	0.442	0.435	0.429	0.461	0.472	0.429	0.447	1	0	0	0	0	0	0
GLR_HUMAN	0.418	0.444	0.410	0.407	0.443	0.438	0.472	0	0	0	2	2	0	2	0.450	0.494	0.421	0.431	0.435	0.439	0.462	0	0	0	2	2	0	0	0.436	0.466	0.446	0.479	0.456	0.430	0.477	0	0	0	0	2	0	1
MGR8_HUMAN	0.386	0.382	0.384	0.364	0.397	0.421	0.399	2	0	0	2	0	0	0	0.422	0.438	0.405	0.388	0.406	0.430	0.410	2	0	0	0	0	0	0	0.441	0.448	0.441	0.423	0.411	0.402	0.461	2	0	0	0	0	0	0
GPC5B_HUMAN	0.397	0.413	0.396	0.394	0.463	0.429	0.397	0	2	0	0	0	0	1	0.442	0.468	0.424	0.454	0.473	0.473	0.437	0	0	0	0	0	0	0	0.450	0.441	0.420	0.478	0.459	0.480	0.457	0	1	0	0	0	0	0
FZD6_HUMAN	0.387	0.368	0.361	0.376	0.394	0.402	0.408	0	0	0	0	0	0	0	0.434	0.415	0.381	0.390	0.400	0.424	0.384	0	0	0	0	0	0	0	0.415	0.415	0.437	0.416	0.401	0.411	0.437	0	0	0	2	1	0	0
CML2_HUMAN	0.437	0.494	0.427	0.377	0.484	0.517	0.516	2	2	2	2	2	2	2	0.520	0.598	0.460	0.433	0.498	0.585	0.441	2	2	2	2	1	2	2	0.514	0.524	0.616	0.480	0.531	0.600	0.635	2	2	2	2	1	2	2
NPBW2_HUMAN	0.410	0.500	0.429	0.391	0.444	0.508	0.523	2	2	2	2	2	2	1	0.437	0.605	0.456	0.427	0.481	0.590	0.462	2	2	2	2	2	2	1	0.523	0.594	0.560	0.464	0.541	0.605	0.610	2	2	2	0	2	2	1
P2Y10_HUMAN	0.376	0.448	0.435	0.402	0.407	0.460	0.460	2	2	1	0	1	2	2	0.420	0.513	0.461	0.442	0.433	0.509	0.390	2	2	1	0	1	2	0	0.453	0.475	0.551	0.439	0.471	0.507	0.474	2	2	1	0	1	2	2
GP158_HUMAN	0.389	0.390	0.364	0.364	0.393	0.406	0.412	2	0	0	1	0	0	0	0.425	0.438	0.390	0.403	0.404	0.421	0.402	1	0	0	1	0	0	0	0.441	0.430	0.402	0.447	0.422	0.451	0.491	1	0	0	0	0	0	0
T2R43_HUMAN	0.395	0.371	0.385	0.382	0.393	0.402	0.388	1	1	0	2	2	2	2	0.395	0.422	0.376	0.373	0.396	0.428	0.374	0	1	0	0	0	2	0	0.397	0.401	0.445	0.406	0.425	0.406	0.403	0	0	0	0	0	1	0
VN1R2_HUMAN	0.374	0.434	0.364	0.364	0.392	0.401	0.406	0	0	0	0	0	0	0	0.418	0.510	0.402	0.406	0.391	0.443	0.397	0	0	0	0	0	0	0	0.417	0.461	0.435	0.437	0.434	0.423	0.471	0	0	0	1	2	0	0
O52A1_HUMAN	0.408	0.378	0.405	0.369	0.420	0.409	0.405	2	0	0	2	0	2	0	0.417	0.445	0.388	0.387	0.419	0.437	0.404	2	0	0	0	0	2	0	0.456	0.418	0.430	0.427	0.409	0.406	0.453	1	0	0	2	0	1	0
Test set 24																																										
Q5VTM0_HUMAN	0.473	0.494	0.386	0.424	0.424	0.437	0.458	1	2	0	2	2	2	2	0.527	0.546	0.427	0.423	0.435	0.457	0.394	1	2	2	2	0	2	2	0.531	0.495	0.462	0.429	0.456	0.435	0.475	1	2	0	0	0	2	2
BRS3_HUMAN	0.398	0.451	0.350	0.391	0.441	0.498	0.415	1	2	0	2	2	2	2	0.465	0.505	0.372	0.462	0.444	0.553	0.419	1	2	0	2	2	2	2	0.475	0.467	0.402	0.489	0.435	0.549	0.490	1	2	0	2	2	2	2
GP173_HUMAN	0.391	0.406	0.378	0.406	0.411	0.454	0.454	2	2	2	2	0	2	2	0.456	0.474	0.402	0.455	0.423	0.487	0.431	2	0	0	2	0	2	0	0.427	0.444	0.422	0.470	0.427	0.495	0.506	2	0	1	2	0	2	2
EDG8_HUMAN	0.425	0.444	0.399	0.407	0.419	0.468	0.514	0	2	0	0	2	2	2	0.472	0.499	0.451	0.460	0.433	0.491	0.425	0	2	0	2	2	2	2	0.474	0.461	0.430	0.499	0.443	0.467	0.562	0	0	0	2	2	2	2
ADA2C_HUMAN	0.458	0.464	0.395	0.437	0.432	0.547	0.539	2	2	0	2	2	2	2	0.555	0.552	0.435	0.486	0.487	0.602	0.469	2	2	0	2	2	2	2	0.514	0.450	0.442	0.472	0.455	0.554	0.602	2	2	0	2	2	2	2
5HT1E_HUMAN	0.398	0.466	0.386	0.426	0.428	0.480	0.502	0	2	2	2	0	2	2	0.423	0.556	0.470	0.469	0.457	0.529	0.400	0	2	2	2	0	2	0	0.423	0.449	0.445	0.511	0.490	0.473	0.501	0	2	2	2	0	2	2
TAAR2_HUMAN	0.363	0.461	0.408	0.372	0.404	0.440	0.446	0	2	0	0	0	2	2	0.404	0.551	0.403	0.399	0.428	0.492	0.417	0	2	0	0	2	2	0	0.431	0.500	0.419	0.420	0.428	0.456	0.463	0	2	0	0	0	2	2
ACM5_HUMAN	0.446	0.449	0.403	0.429	0.432	0.460	0.498	2	2	0	2	2	2	2	0.497	0.491	0.438	0.434	0.498	0.504	0.441	2	2	2	2	2	2	0	0.478	0.452	0.456	0.474	0.465	0.504	0.530	2	2	0	0	2	2	2
CXCR6_HUMAN	0.436	0.490	0.420	0.428	0.468	0.443	0.473	2	2	2	2	2	2	2	0.496	0.575	0.477	0.491	0.467	0.506	0.388	2	2	2	2	2	2	1	0.571	0.567	0.522	0.580	0.518	0.541	0.555	2	2	2	1	2	2	2
OPRX_HUMAN	0.448	0.490	0.433	0.429	0.496	0.484	0.571	2	2	2	2	2	2	2	0.484	0.594	0.475	0.465	0.505	0.549	0.490	2	2	2	2	2	2	2	0.566	0.583	0.544	0.471	0.520	0.565	0.685	2	2	2	2	2	2	2
C5ARL_HUMAN	0.475	0.476	0.450	0.412	0.423	0.516	0.486	2	2	1	2	2	2	2	0.546	0.581	0.473	0.452	0.461	0.568	0.449	2	2	2	2	2	2	2	0.559	0.562	0.507	0.541	0.525	0.566	0.640	2	2	2	2	2	1	2

Continued

P2Y12_HUMAN	0.422	0.415	0.383	0.396	0.413	0.441	0.461	1	0	0	2	0	2	2	0.441	0.455	0.388	0.429	0.425	0.486	0.407	1	0	0	2	0	1	2	0.521	0.462	0.456	0.454	0.416	0.504	0.473	1	2	2	2	0	2	2
PAR4_HUMAN	0.439	0.481	0.387	0.391	0.437	0.421	0.529	2	2	0	0	2	2	2	0.509	0.565	0.466	0.427	0.467	0.439	0.457	2	2	2	0	2	2	2	0.548	0.542	0.443	0.466	0.507	0.447	0.573	2	1	2	0	2	2	2
EMR1_HUMAN	0.392	0.401	0.402	0.402	0.417	0.430	0.444	0	0	0	0	0	0	1	0.446	0.473	0.414	0.421	0.419	0.439	0.418	0	0	0	0	0	0	0	0.442	0.435	0.429	0.461	0.472	0.429	0.447	1	0	0	0	0	0	0
MGR8_HUMAN	0.386	0.382	0.384	0.364	0.397	0.421	0.399	2	0	0	2	0	0	0	0.422	0.438	0.405	0.388	0.406	0.430	0.410	2	0	0	0	0	0	0	0.441	0.448	0.441	0.423	0.411	0.402	0.461	2	0	0	0	0	0	0
FZD6_HUMAN	0.387	0.368	0.361	0.376	0.394	0.402	0.408	0	0	0	0	0	0	0	0.434	0.415	0.381	0.390	0.400	0.424	0.384	0	0	0	0	0	0	0	0.415	0.415	0.437	0.416	0.401	0.411	0.437	0	0	0	2	1	0	0
NPBW2_HUMAN	0.410	0.500	0.429	0.391	0.444	0.508	0.523	2	2	2	2	2	2	1	0.437	0.605	0.456	0.427	0.481	0.590	0.462	2	2	2	2	2	2	1	0.523	0.594	0.560	0.464	0.541	0.605	0.610	2	2	2	0	2	2	1
VN1R2_HUMAN	0.374	0.434	0.364	0.364	0.392	0.401	0.406	0	0	0	0	0	0	0	0.418	0.510	0.402	0.406	0.391	0.443	0.397	0	0	0	0	0	0	0	0.417	0.461	0.435	0.437	0.434	0.423	0.471	0	0	0	1	2	0	0
OR2C1_HUMAN	0.398	0.412	0.392	0.391	0.405	0.413	0.454	0	2	0	0	2	2	2	0.448	0.468	0.396	0.416	0.423	0.440	0.407	0	0	0	0	2	2	0	0.465	0.455	0.450	0.484	0.422	0.437	0.462	0	0	0	2	0	2	2
AGTR1_BOVIN	0.453	0.418	0.388	0.373	0.435	0.474	0.479	2	2	2	0	2	2	1	0.520	0.481	0.411	0.388	0.446	0.529	0.438	2	2	2	2	2	2	1	0.557	0.515	0.501	0.484	0.493	0.549	0.549	2	2	2	2	2	2	1
Q3T181_BOVIN	0.394	0.475	0.384	0.427	0.452	0.517	0.510	0	2	0	1	0	2	2	0.429	0.541	0.456	0.502	0.486	0.553	0.466	0	2	2	2	1	2	2	0.486	0.535	0.532	0.533	0.534	0.557	0.616	2	2	2	2	1	2	2
FSHR_BOVIN	0.415	0.467	0.388	0.412	0.415	0.426	0.480	2	2	0	2	0	0	2	0.477	0.526	0.425	0.424	0.425	0.460	0.412	2	2	2	2	0	2	2	0.469	0.499	0.478	0.447	0.432	0.435	0.567	2	2	0	2	2	0	2
GNRHR_CANFA	0.374	0.437	0.409	0.386	0.451	0.429	0.434	0	2	2	1	2	1	2	0.425	0.482	0.461	0.409	0.467	0.472	0.397	0	2	2	0	2	1	0	0.412	0.440	0.467	0.437	0.444	0.475	0.474	0	2	0	1	2	1	0
EDNRA_CANFA	0.392	0.460	0.364	0.423	0.404	0.475	0.401	2	2	0	2	0	2	2	0.465	0.491	0.428	0.469	0.429	0.509	0.387	2	2	2	2	0	2	0	0.472	0.449	0.400	0.513	0.409	0.490	0.434	1	1	0	2	2	2	2
5HT1A_CANFA	0.400	0.465	0.416	0.405	0.454	0.544	0.552	2	2	2	2	2	2	2	0.455	0.560	0.452	0.440	0.479	0.610	0.473	2	2	2	2	0	2	2	0.454	0.456	0.452	0.476	0.470	0.533	0.593	2	2	0	0	0	2	2
OPS5_DROME	0.434	0.485	0.369	0.374	0.426	0.417	0.390	2	2	0	0	0	2	0	0.478	0.560	0.385	0.413	0.439	0.477	0.401	2	2	0	0	0	2	0	0.470	0.519	0.407	0.474	0.442	0.481	0.419	2	2	0	2	2	2	0
GR33A_DROME	0.410	0.396	0.371	0.385	0.403	0.414	0.445	0	0	0	0	1	0	0	0.459	0.468	0.392	0.400	0.396	0.420	0.424	0	0	0	0	1	0	0	0.433	0.435	0.424	0.411	0.412	0.441	0.500	0	0	0	0	1	1	0
MTH4_DROME	0.369	0.407	0.401	0.375	0.400	0.394	0.397	0	0	2	1	0	0	1	0.401	0.471	0.424	0.398	0.426	0.416	0.395	0	0	2	2	2	0	0	0.435	0.428	0.450	0.426	0.415	0.407	0.428	0	0	0	2	0	0	0
OLF6_CHICK	0.415	0.404	0.373	0.373	0.400	0.398	0.444	0	0	0	0	2	0	2	0.401	0.467	0.400	0.380	0.415	0.426	0.412	0	0	0	0	0	0	0	0.419	0.418	0.459	0.402	0.421	0.413	0.454	0	0	0	0	0	0	2
FZD7_CHICK	0.392	0.453	0.422	0.389	0.413	0.410	0.414	0	0	0	0	0	1	0	0.428	0.510	0.429	0.430	0.426	0.438	0.432	0	0	2	0	1	0	0	0.429	0.448	0.446	0.467	0.465	0.435	0.461	0	0	2	0	1	1	0
P2RY1_CHICK	0.453	0.514	0.437	0.407	0.448	0.493	0.551	2	2	2	0	2	2	2	0.509	0.592	0.468	0.450	0.482	0.546	0.478	2	2	2	2	2	2	2	0.562	0.592	0.569	0.473	0.451	0.559	0.646	2	2	2	2	2	1	2
TAAR6_MOUSE	0.441	0.442	0.369	0.414	0.405	0.477	0.477	1	2	1	2	2	2	1	0.458	0.473	0.432	0.451	0.442	0.526	0.401	1	2	2	2	2	1	0	0.498	0.435	0.416	0.459	0.433	0.507	0.494	1	0	2	0	1	2	1
NMBR_MOUSE	0.461	0.460	0.373	0.417	0.471	0.507	0.415	2	2	2	2	2	2	2	0.521	0.554	0.427	0.488	0.504	0.578	0.421	2	2	2	2	2	2	2	0.552	0.511	0.426	0.530	0.486	0.567	0.466	1	2	0	2	2	2	2
FFAR2_MOUSE	0.423	0.478	0.393	0.390	0.431	0.449	0.452	2	2	0	2	2	2	2	0.490	0.570	0.438	0.430	0.465	0.474	0.421	2	2	1	2	2	1	0	0.507	0.565	0.503	0.454	0.455	0.466	0.474	1	2	1	0	0	2	0
MLO10_ARATH	0.384	0.385	0.385	0.380	0.405	0.406	0.388	2	0	0	0	0	0	0	0.414	0.416	0.402	0.446	0.442	0.421	0.398	1	0	0	0	0	0	0	0.397	0.419	0.424	0.496	0.424	0.417	0.442	2	0	0	0	2	0	0
5HT2A_MACMU	0.389	0.453	0.413	0.409	0.462	0.442	0.454	2	2	2	0	2	2	2	0.437	0.519	0.464	0.395	0.484	0.485	0.390	2	2	2	0	2	2	0	0.446	0.460	0.426	0.426	0.463	0.449	0.457	0	2	0	0	2	2	2
OPSD_MESBI	0.402	0.505	0.410	0.424	0.492	0.506	0.457	2	2	0	0	2	2	2	0.453	0.597	0.421	0.468	0.514	0.543	0.438	0	2	0	2	2	2	0	0.435	0.561	0.464	0.462	0.492	0.570	0.498	2	2	0	0	2	2	2
V1AR_MICOH	0.449	0.432	0.413	0.391	0.426	0.490	0.490	2	1	2	0	1	2	2	0.509	0.474	0.470	0.417	0.452	0.574	0.453	2	2	1	0	1	2	2	0.534	0.440	0.461	0.443	0.465	0.536	0.589	2	2	1	0	1	2	2

Note: The number "0", "1", and "2" donate that the predicting helical regions of GPCRs by the scoring matrices of the training set 18 $\left(S_{D18}^{L}\right)$ are "different from", "partial consistent with (similar to)", and "identical with" the actual transmembrane regions of GPCRs, respectively.

3. RESULTS

In what follows, we present three primary results, based on application of the methods described above.

3.1. Phylogenetic Analysis and Structural Evolution

Figure 1 displays the grouping frame of the training datasets (learning dataset, S^{L}), where 22 groups belong to three types. Of the different chromosome type, there are five classes: Class A (Groups 1-4 and Group 14), Class B (Group 15), Class C (Group 16), Class F (Group 22), and Class O (Groups 5-13 and Group 17). Group 1 $\left(S_{A1}^{L}\right)$, Group 2 $\left(S_{A2}^{L}\right)$, Group 3 $\left(S_{A3}^{L}\right)$, Group 4 $\left(S_{A4}^{L}\right)$, and Group 14 $\left(S_{A14}^{L}\right)$ of Class A are composed of 44, 38, 32, 20, and 44 GPCRs, respectively. Class B $\left(S_{B15}^{L}\right)$, C $\left(S_{C16}^{L}\right)$, and F/S $\left(S_{F22}^{L}\right)$ contain 13, 10, and 9 GPCRs, respectively. Groups $5,6,\cdots$ and 13 $\left(S_{O5}^{L},S_{O6}^{L},\cdots,S_{O13}^{L}\right)$ consist of 39, 24, 33, 11, 48, 40, 44, 20, and 20 GPCRs, respectively; while Group 17 $\left(S_{O17}^{L}\right)$ includes 33 GPCRs. Group 18 $\left(S_{D18}^{L}\right)$, Group 19 $\left(S_{S19}^{L}\right)$,Group 20 $\left(S_{S20}^{L}\right)$, and Group 21 $\left(S_{D21}^{L}\right)$ contain 39, 27, 22, and 20 GPCRs, respectively. The following test datasets $\left(S^{T}\right)$ are composed of two groups: Group 23 $\left(S_{23}^{T}\right)$ from human and Group 24 $\left(S_{24}^{T}\right)$ from different species (Appendix 1, **Table 3**). Here, $S_{A1}^{L}\bigcap S_{A14}^{L}\neq\Phi$, $S_{A2}^{L}\bigcap S_{A14}^{L}\neq\Phi$, $S_{A3}^{L}\bigcap S_{A14}^{L}\neq\Phi$, and $S_{A4}^{L}\bigcap S_{A14}^{L}\neq\Phi$; $S_{O5}^{L}\bigcap S_{O17}^{L}\neq\Phi$, $S_{O6}^{L}\bigcap S_{O17}^{L}\neq\Phi$, $S_{O7}^{L}\bigcap S_{O17}^{L}\neq\Phi$, $S_{O8}^{L}\bigcap S_{O17}^{L}\neq\Phi$, $S_{O9}^{L}\bigcap S_{O17}^{L}\neq\Phi$, $S_{O10}^{L}\bigcap S_{O17}^{L}\neq\Phi$, $S_{O11}^{L}\bigcap S_{O17}^{L}\neq\Phi$, $S_{O12}^{L}\bigcap S_{O17}^{L}\neq\Phi$ and $S_{O13}^{L}\bigcap S_{O17}^{L}\neq\Phi$; $S_{D18}^{L}\bigcap S_{A14}^{L}\neq\Phi$, $S_{D18}^{L}\bigcap S_{B15}^{L}\neq\Phi$, $S_{D18}^{L}\bigcap S_{C16}^{L}\neq\Phi$, $S_{D18}^{L}\bigcap S_{F22}^{L}\neq\Phi$ and $S_{D18}^{L}\bigcap S_{O17}^{L}\neq\Phi$.

Table 1 lists the amino acid sequences of TM1 in Group 1 GPCRs, the common 12-residue regions of TM1 by alignment, and the corresponding coding sequences consisting of 36 nucleotides. **Table 2** displays the amino acid sequence length and the sample number consisting of the scoring matrix of each transmembrane region of GPCRs in the training datasets after sequence alignments. Different the training sets, different the amino acid sequence length and the sample number consisting of the scoring matrix to same TMs; the same the training sets, different the amino acid sequence length and the sample number consisting of the scoring matrix to different TMs. **Figure 2** illustrates the scoring matrices of seven TMs (TM1-TM7) of GPCRs in Group 1 of Class A in the training datasets. This is the core of prediction system of GPCRs.

3.2. Validation of the Models (Scoring Matrix)

22 scoring matrices are built based on the 22 groups of training datasets $\left(S^L\right)$ from GPCR superfamily and validated by two groups of test sets $\left(S^T\right)$. **Tables 4** and **5** display the score and the prediction accuracy of the coding sequences of GPCRs' trans-membrane segments in test sets by the scoring matrices of the different training datasets. All the data can be clearly divided into four categories: S_{23}^T-"2", S_{23}^T-"1 + 2", S_{24}^T-"2" and S_{24}^T-"1 + 2". The number "2" donates that the predicting helical regions of GPCRs by the scoring matrices of the training set are identical with the actual TM regions of GPCRs, while the number "1" donates the predicting helical regions of GPCRs are partial consistent with their actual TMs. "1 + 2" means the combination of "2" with "1", namely the positive prediction results. There are 22 examinations (corresponding 22 training sets) in each category. For instance, if we use test set S_{23}^T to examine Group 1 $\left(S_{A1}^L\right)$ of training set, and then get seven "all hit" ("2") correctness rates (validity) (TM1 to TM7). These 7 correctness rates as a whole can be deemed as examination 1 in S_{23}^T-"2". Under this situation, mean of one examination's correctness rate is the mean of the seven correctness rates (**Table 5**). The rest may be deduced by analogy.

3.3. Statistics Analysis

One way-ANOVA, a powerful and common statistical procedure, is used to figure out whether there are significant differences among means of correctness rate of the examinations. As one way-ANOVA requires, all data that does not obey normal distribution are eliminated, such as Groups 1, 6 and 10 of S_{23}^T-"2", group 16 of S_{23}^T-"1 + 2", groups 12, 16 and 22 of S_{24}^T-"2", and groups 15, 19 and 22 of S_{24}^T-"1 + 2". The four results of one way-ANOVA, with F values of 21.931, 9.308, 22.807 and 7.488 for S_{23}^T-"2", S_{23}^T-"1 + 2", S_{24}^T-"2" and S_{24}^T-"1 + 2", respectively, indicate that there are significant differences between means of correctness rate of examinations at the 0.01 level in each category. Then Test of Homogeneity of Variances is applied in order to find a suitable method for multiple comparisons. Actually, S_{23}^T-"1 + 2" and S_{24}^T-"1 + 2", with P values of 0.7673 and 0.7121, respectively, has homogenous variances at the 0.05 level, and LSD method of multiple comparisons will be used. On the other hand, variances of S_{23}^T-"2" and S_{24}^T-"2" with P values of 0.0032 and 0.0418, respectively, are not homogenous, which means that Tamhane's T2 should be chosen as multiple comparisons method. The results of multiple comparisons are respectively visualized in **Figure 4**, where "X" shows there are significant differences (at the significant level of 0.05) between the two examinations indicated by corresponding column and line (specific value of multiple comparisons can be found in supplemental data). Finally, the average of correctness rate of examinations of each category are plotted respectively on **Figure 5**, from which we can see that the training set of 18 has the highest mean of correctness rate when examined by test set in each of the four groups although training set 18 and other training sets such as 2, 3, 4, 14, etc. has not statistical difference. ANOVA results reveal that there are three scoring matrices significant, from three training datasets, Group 3 $\left(S_{A3}^L\right)$, 14 $\left(S_{A14}^L\right)$, and 18 $\left(S_{D18}^L\right)$, respectively (**Figure 5**).

The following t-test results reveal that the mean difference of the scores between two groups, "2"/"2 + 1" and "0", are statistically significant ($P < 0.05$) with the exception of TM4 and TM5 (**Table 6**) based on the scores and the validity of GPCRs in test sets (S_{23}^T and S_{24}^T) by the scoring matrices of training set S_{D18}^L, S_{A14}^L and S_{A3}^L (**Table 3**). Especially to TM2, TM3, TM6, and TM7, the mean of two groups between "2"/"2 + 1" and "0" is statistically significant ($P < 0.01$), which means that the probability of the difference being due to chance is less than 0.01. Of the seven TMs, there is significant different between the scores of two groups in TM2 with the t values from 4.494 to 6.959 ($P < 0.001$), where the degrees of freedom of a set of data are more than 30 but less than 40 (the critical value of t for the 0.001 level of significance at 30 of df is 3.646). On the other hand, t-values show that there are significant different between the scores of two groups by the scoring matrix of S_{A14}^L ($P < 0.001$) except in TM1, TM6, and TM7 of S_{23}^T and TM5 of S_{24}^T ($P < 0.01$). But the probability that the difference between samples in TM4 of S_{D18}^L and TM5 of S_{A3}^L is more than 0.05 due to sampling error.

Comparison of t-values in S_{23}^T with those in S_{24}^T reveals that different scoring matrices have different statistical significance. To the scoring matrix of S_{D18}^L and S_{A14}^L, the mean differences between the scores of two groups in S_{24}^T are more than those in S_{23}^T; whereas the mean differences in S_{23}^T are more than those in S_{24}^T. The reason may be the homology of samples consisting of training sets and test sets. The samples of S_{D18}^L and S_{A14}^L come from different GPCR sub-families whereas those of S_{A3}^L come from the same GPCR subfamily. Similarly, the members of S_{23}^T belong to human GPCR proteins whereas those of S_{24}^T are GPCR proteins from different species.

Statistical graphs reveal that the mean scores of the predicting coding sequences of GPCRs' 7 TMs in test sets (S_{23}^T and S_{24}^T) by the scoring matrices of S_{A3}^L is higher than those by S_{A14}^L except TM2 and TM6, while the predicting scores by S_{A14}^L are higher than those by S_{D18}^L except TM7 (**Figure 6**). **Figure 7** shows the histo-

Table 4. The score of the coding sequences of GPCRs' each trans-membrane segment in test sets by the scoring matrices of the different training datasets.

Training set	Test set 23							Test set 24						
	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7
"2"*														
1	0.445 ± 0.035	0.481 ± 0.033	0.416 ± 0.021	0.426 ± 0.030	0.457 ± 0.031	0.520 ± 0.034	0.473 ± 0.039	0.454 ± 0.034	0.490 ± 0.027	0.426 ± 0.024	0.434 ± 0.033	0.470 ± 0.031	0.531 ± 0.041	0.480 ± 0.034
2	0.461 ± 0.055	0.527 ± 0.064	0.509 ± 0.064	0.481 ± 0.038	0.494 ± 0.040	0.520 ± 0.048	0.536 ± 0.066	0.460 ± 0.049	0.538 ± 0.061	0.504 ± 0.077	0.484 ± 0.038	0.508 ± 0.044	0.534 ± 0.056	0.538 ± 0.068
3	0.503 ± 0.052	0.519 ± 0.054	0.508 ± 0.058	0.475 ± 0.033	0.478 ± 0.037	0.500 ± 0.049	0.550 ± 0.070	0.500 ± 0.056	0.508 ± 0.053	0.495 ± 0.052	0.483 ± 0.035	0.469 ± 0.038	0.509 ± 0.049	0.535 ± 0.072
4	0.469 ± 0.030	0.487 ± 0.059	0.484 ± 0.032	0.444 ± 0.021	0.436 ± 0.031	0.513 ± 0.053	0.501 ± 0.057	0.470 ± 0.035	0.482 ± 0.055	0.475 ± 0.037	0.445 ± 0.024	0.440 ± 0.024	0.515 ± 0.036	0.500 ± 0.057
5	0.417 ± 0.012	0.455 ± 0.012	N/A	N/A	0.447 ± 0.052	0.431 ± 0.012	N/A	0.427 ± 0.018	N/A	N/A	N/A	0.424 ± 0.016	0.430 ± 0.014	N/A
6	0.518 ± 0.034	0.477 ± 0.034	N/A	0.426 ± 0.019	0.413 ± 0.009	0.420 ± 0.014	0.425 ± 0.022	0.500 ± 0.025	0.462 ± 0.029	0.488 ± 0.045	0.428 ± 0.018	0.410 ± 0.005	0.460 ± 0.097	0.444 ± 0.031
7	0.439 ± 0.017	0.460 ± 0.021	N/A	N/A	0.419 ± 0.017	N/A	0.462 ± 0.017	0.457 ± 0.046	0.523 ± 0.098	N/A	0.382 ± 0.036	0.418 ± 0.013	N/A	0.501 ± 0.064
8	0.452 ± 0.022	N/A	N/A	0.388 ± 0.002	0.408 ± 0.018	N/A	0.449 ± 0.018	0.459 ± 0.040	0.626 ± 0.105	N/A	N/A	0.404 ± 0.012	0.501 ± 0.102	0.445 ± 0.045
9	0.461 ± 0.029	0.431 ± 0.032	0.504 ± 0.030	0.414 ± 0.015	0.447 ± 0.034	0.453 ± 0.019	0.449 ± 0.032	0.468 ± 0.030	0.453 ± 0.061	0.592 ± 0.115	0.416 ± 0.015	0.459 ± 0.034	0.478 ± 0.062	0.452 ± 0.030
10	0.422 ± 0.006	N/A	N/A	N/A	N/A	N/A	0.413 ± 0.025	0.414 ± 0.017	N/A	N/A	0.408 ± 0.001	N/A	N/A	0.435 ± 0.030
11	0.457 ± 0.029	0.483 ± 0.033	0.404 ± 0.041	0.400 ± 0.001	0.405 ± 0.018	N/A	0.426 ± 0.033	0.475 ± 0.038	0.537 ± 0.077	0.449 ± 0.073	0.421 ± 0.023	0.407 ± 0.023	0.600 ± 0.097	0.442 ± 0.029
12	0.432 ± 0.017	N/A	0.453 ± 0.049	0.416 ± 0.015	0.412 ± 0.002	N/A	0.421 ± 0.022	0.454 ± 0.048	0.595 ± 0.088	N/A	0.415 ± 0.022	0.402 ± 0.016	0.582 ± 0.050	0.430 ± 0.056
13	0.467 ± 0.025	0.471 ± 0.034	N/A	N/A	0.385 ± 0.012	N/A	0.416 ± 0.025	0.480 ± 0.040	0.529 ± 0.051	N/A	0.447 ± 0.047	0.393 ± 0.015	0.459 ± 0.107	N/A
14	0.479 ± 0.045	0.548 ± 0.040	0.454 ± 0.021	0.447 ± 0.023	0.464 ± 0.025	0.509 ± 0.049	0.443 ± 0.027	0.488 ± 0.036	0.540 ± 0.042	0.448 ± 0.021	0.450 ± 0.029	0.466 ± 0.028	0.523 ± 0.050	0.444 ± 0.030
15	0.510 ± 0.067	0.409 ± 0.009	0.503 ± 0.086	0.512 ± 0.083	0.496 ± 0.042	0.444 ± 0.039	0.497 ± 0.079	0.482 ± 0.053	0.446 ± 0.036	0.531 ± 0.082	0.489 ± 0.038	0.415 ± 0.035	0.419 ± 0.016	0.484 ± 0.069
16	0.474 ± 0.033	N/A	0.479 ± 0.063	0.482 ± 0.066	0.477 ± 0.015	0.583 ± 0.015	0.467 ± 0.089	0.448 ± 0.051	N/A	0.451 ± 0.032	N/A	0.472 ± 0.030	0.597 ± 0.129	N/A
17	0.458 ± 0.032	0.432 ± 0.019	0.442 ± 0.049	0.412 ± 0.020	0.420 ± 0.022	N/A	0.420 ± 0.028	0.465 ± 0.029	0.527 ± 0.106	0.471 ± 0.058	0.406 ± 0.019	0.427 ± 0.018	N/A	0.430 ± 0.027
18	0.421 ± 0.029	0.465 ± 0.027	0.410 ± 0.020	0.402 ± 0.024	0.438 ± 0.030	0.464 ± 0.036	0.481 ± 0.046	0.426 ± 0.029	0.465 ± 0.027	0.407 ± 0.021	0.411 ± 0.019	0.441 ± 0.026	0.474 ± 0.037	0.478 ± 0.045
19	0.437 ± 0.029	0.463 ± 0.031	0.417 ± 0.014	0.410 ± 0.026	0.443 ± 0.031	0.460 ± 0.031	0.458 ± 0.041	0.437 ± 0.031	0.460 ± 0.035	0.424 ± 0.02	0.414 ± 0.02	0.440 ± 0.032	0.467 ± 0.038	0.456 ± 0.040
20	0.464 ± 0.042	0.521 ± 0.039	0.394 ± 0.010	0.392 ± 0.021	0.435 ± 0.027	0.403 ± 0.016	0.447 ± 0.025	0.466 ± 0.040	0.518 ± 0.031	0.405 ± 0.014	0.392 ± 0.018	0.428 ± 0.027	0.408 ± 0.021	0.457 ± 0.033
21	0.410 ± 0.020	0.444 ± 0.020	0.381 ± 0.014	0.387 ± 0.018	0.425 ± 0.021	0.436 ± 0.028	0.415 ± 0.024	0.418 ± 0.023	0.446 ± 0.023	0.385 ± 0.022	0.390 ± 0.020	0.435 ± 0.026	0.439 ± 0.028	0.421 ± 0.029
22	0.469 ± 0.023	N/A	0.565 ± 0.151	N/A	N/A	N/A	N/A	0.508 ± 0.083	N/A	0.580 ± 0.129	0.638 ± 0.177	0.622 ± 0.168	N/A	N/A
PreMod	0.421 ± 0.029	0.465 ± 0.027	0.454 ± 0.021	0.447 ± 0.023	0.438 ± 0.030	0.464 ± 0.036	0.481 ± 0.046	0.426 ± 0.029	0.465 ± 0.027	0.448 ± 0.021	0.450 ± 0.029	0.441 ± 0.026	0.474 ± 0.037	0.478 ± 0.045
"1 + 2"*														
1	0.442 ± 0.034	0.477 ± 0.037	0.413 ± 0.023	0.427 ± 0.030	0.457 ± 0.030	0.515 ± 0.036	0.473 ± 0.041	0.454 ± 0.033	0.489 ± 0.027	0.423 ± 0.026	0.434 ± 0.034	0.470 ± 0.031	0.524 ± 0.041	0.475 ± 0.036
2	0.454 ± 0.049	0.523 ± 0.066	0.495 ± 0.062	0.479 ± 0.038	0.495 ± 0.039	0.516 ± 0.048	0.537 ± 0.069	0.456 ± 0.046	0.525 ± 0.064	0.493 ± 0.077	0.480 ± 0.037	0.508 ± 0.044	0.525 ± 0.057	0.542 ± 0.066
3	0.494 ± 0.049	0.517 ± 0.054	0.505 ± 0.059	0.478 ± 0.039	0.477 ± 0.040	0.496 ± 0.053	0.552 ± 0.071	0.500 ± 0.049	0.507 ± 0.053	0.488 ± 0.050	0.483 ± 0.041	0.465 ± 0.041	0.507 ± 0.050	0.537 ± 0.069
4	0.468 ± 0.029	0.485 ± 0.057	0.470 ± 0.043	0.448 ± 0.025	0.436 ± 0.031	0.506 ± 0.055	0.499 ± 0.055	0.464 ± 0.036	0.477 ± 0.053	0.467 ± 0.042	0.449 ± 0.027	0.440 ± 0.022	0.508 ± 0.040	0.499 ± 0.055
5	0.418 ± 0.052	0.461 ± 0.045	0.449 ± 0.065	0.376 ± 0.026	0.437 ± 0.043	0.440 ± 0.029	0.425 ± 0.049	0.418 ± 0.017	0.469 ± 0.041	0.439 ± 0.026	0.369 ± 0.012	0.422 ± 0.020	0.430 ± 0.014	0.422 ± 0.037
6	0.470 ± 0.035	0.464 ± 0.030	0.448 ± 0.035	0.425 ± 0.018	0.400 ± 0.017	0.420 ± 0.012	0.438 ± 0.027	0.469 ± 0.021	0.459 ± 0.040	0.475 ± 0.047	0.426 ± 0.019	0.401 ± 0.021	0.473 ± 0.081	0.465 ± 0.075
7	0.443 ± 0.018	0.469 ± 0.031	0.448 ± 0.034	0.367 ± 0.002	0.425 ± 0.020	0.446 ± 0.045	0.454 ± 0.019	0.449 ± 0.032	0.472 ± 0.056	0.472 ± 0.078	0.382 ± 0.026	0.423 ± 0.014	0.466 ± 0.085	0.476 ± 0.053
8	0.430 ± 0.025	0.486 ± 0.029	0.443 ± 0.030	0.399 ± 0.029	0.405 ± 0.014	0.438 ± 0.026	0.447 ± 0.022	0.443 ± 0.033	0.494 ± 0.067	0.503 ± 0.066	0.401 ± 0.025	0.406 ± 0.009	0.470 ± 0.082	0.466 ± 0.075
9	0.455 ± 0.029	0.439 ± 0.029	0.500 ± 0.024	0.412 ± 0.014	0.443 ± 0.033	0.451 ± 0.020	0.451 ± 0.028	0.463 ± 0.031	0.450 ± 0.052	0.563 ± 0.110	0.418 ± 0.021	0.455 ± 0.032	0.458 ± 0.053	0.470 ± 0.069
10	0.414 ± 0.022	0.429 ± 0.025	0.428 ± 0.033	0.388 ± 0.015	0.388 ± 0.018	0.421 ± 0.033	0.423 ± 0.027	0.422 ± 0.025	0.437 ± 0.054	0.445 ± 0.056	0.396 ± 0.029	0.393 ± 0.020	0.473 ± 0.096	0.448 ± 0.062
11	0.453 ± 0.031	0.480 ± 0.031	0.398 ± 0.018	0.402 ± 0.018	0.405 ± 0.018	0.429 ± 0.011	0.436 ± 0.026	0.457 ± 0.039	0.494 ± 0.052	0.420 ± 0.053	0.413 ± 0.021	0.410 ± 0.020	0.508 ± 0.118	0.459 ± 0.071
12	0.422 ± 0.018	0.445 ± 0.026	0.426 ± 0.027	0.418 ± 0.014	0.404 ± 0.018	0.442 ± 0.044	0.418 ± 0.025	0.435 ± 0.039	0.457 ± 0.064	0.452 ± 0.073	0.410 ± 0.023	0.401 ± 0.019	0.526 ± 0.102	0.442 ± 0.054
13	0.465 ± 0.022	0.473 ± 0.031	0.418 ± 0.021	0.398 ± 0.009	0.392 ± 0.022	0.398 ± 0.020	0.430 ± 0.026	0.474 ± 0.035	0.502 ± 0.052	0.465 ± 0.074	0.419 ± 0.038	0.404 ± 0.030	0.446 ± 0.088	0.474 ± 0.085
14	0.477 ± 0.045	0.543 ± 0.046	0.455 ± 0.020	0.446 ± 0.024	0.464 ± 0.026	0.507 ± 0.047	0.445 ± 0.035	0.482 ± 0.038	0.540 ± 0.042	0.449 ± 0.021	0.450 ± 0.029	0.462 ± 0.031	0.518 ± 0.048	0.441 ± 0.031
15	0.510 ± 0.067	0.446 ± 0.045	0.497 ± 0.080	0.497 ± 0.074	0.448 ± 0.050	0.442 ± 0.037	0.489 ± 0.075	0.482 ± 0.053	0.442 ± 0.033	0.513 ± 0.080	0.490 ± 0.038	0.424 ± 0.029	0.438 ± 0.032	0.484 ± 0.054
16	0.450 ± 0.034	0.474 ± 0.017	0.479 ± 0.073	0.415 ± 0.070	0.496 ± 0.049	0.475 ± 0.095	0.455 ± 0.060	0.439 ± 0.033	0.472 ± 0.024	0.461 ± 0.063	0.432 ± 0.085	0.490 ± 0.052	0.474 ± 0.102	0.473 ± 0.069
17	0.453 ± 0.030	0.440 ± 0.031	0.413 ± 0.027	0.405 ± 0.022	0.413 ± 0.023	0.419 ± 0.027	0.422 ± 0.032	0.462 ± 0.030	0.451 ± 0.056	0.431 ± 0.051	0.403 ± 0.023	0.424 ± 0.022	0.456 ± 0.082	0.440 ± 0.061
18	0.421 ± 0.029	0.461 ± 0.032	0.414 ± 0.022	0.401 ± 0.024	0.437 ± 0.030	0.464 ± 0.036	0.481 ± 0.048	0.427 ± 0.029	0.463 ± 0.027	0.408 ± 0.025	0.409 ± 0.020	0.439 ± 0.026	0.470 ± 0.038	0.476 ± 0.045
19	0.429 ± 0.030	0.463 ± 0.031	0.416 ± 0.015	0.412 ± 0.027	0.439 ± 0.030	0.458 ± 0.036	0.459 ± 0.042	0.436 ± 0.029	0.459 ± 0.035	0.419 ± 0.022	0.418 ± 0.025	0.443 ± 0.030	0.467 ± 0.038	0.454 ± 0.039
20	0.462 ± 0.04	0.512 ± 0.047	0.393 ± 0.016	0.387 ± 0.020	0.437 ± 0.029	0.400 ± 0.016	0.445 ± 0.025	0.464 ± 0.041	0.516 ± 0.032	0.401 ± 0.018	0.391 ± 0.019	0.428 ± 0.025	0.408 ± 0.021	0.451 ± 0.032
21	0.412 ± 0.022	0.437 ± 0.028	0.378 ± 0.014	0.395 ± 0.020	0.423 ± 0.021	0.433 ± 0.030	0.414 ± 0.023	0.417 ± 0.024	0.446 ± 0.023	0.384 ± 0.022	0.396 ± 0.020	0.431 ± 0.027	0.438 ± 0.027	0.419 ± 0.030
22	0.472 ± 0.031	0.423 ± 0.045	0.501 ± 0.073	0.467 ± 0.047	0.487 ± 0.026	0.462 ± 0.033	0.495 ± 0.083	0.492 ± 0.071	0.494 ± 0.101	0.573 ± 0.154	0.503 ± 0.109	0.520 ± 0.127	0.486 ± 0.110	0.550 ± 0.149
PreMod	0.421 ± 0.029	0.461 ± 0.032	0.455 ± 0.020	0.446 ± 0.024	0.437 ± 0.030	0.464 ± 0.036	0.481 ± 0.048	0.427 ± 0.029	0.463 ± 0.027	0.449 ± 0.021	0.450 ± 0.029	0.439 ± 0.026	0.470 ± 0.038	0.476 ± 0.045

Note: *The number "1" and "2" donate that the predicting helical regions of GPCRs by the scoring matrices of the training sets are "partial consistent with" and "identical with" the actual TM regions of GPCRs, respectively.

Table 5. The validity of prediction the coding sequences of GPCRs' each TM segment in test sets by the scoring matrices of the different training sets.

Training set	Test set 23																	Test set 24														
	"2"								"1 + 2"									"2"								"1 + 2"							
	TM1	TM2	TM3	TM4	TM5	TM6	TM7	Mean	TM1	TM2	TM3	TM4	TM5	TM6	TM7	Mean	TM1	TM2	TM3	TM4	TM5	TM6	TM7	Mean	TM1	TM2	TM3	TM4	TM5	TM6	TM7	Mean
1	0.579	0.632	0.211	0.632	0.553	0.605	0.737		0.684	0.658	0.237	0.658	0.579	0.684	0.842	0.620	0.632	0.632	0.316	0.553	0.553	0.579	0.684	0.564	0.658	0.658	0.342	0.684	0.553	0.684	0.816	0.628
2	0.500	0.421	0.263	0.474	0.395	0.632	0.632	0.474	0.711	0.526	0.342	0.500	0.421	0.737	0.816	0.579	0.526	0.474	0.316	0.474	0.474	0.579	0.605	0.493	0.684	0.579	0.368	0.553	0.474	0.763	0.711	0.590
3	0.500	0.553	0.368	0.553	0.553	0.737	0.632	0.557	0.737	0.605	0.421	0.605	0.632	0.816	0.763	0.654	0.474	0.605	0.316	0.474	0.526	0.658	0.632	0.526	0.684	0.658	0.395	0.553	0.684	0.789	0.711	0.639
4	0.500	0.447	0.289	0.500	0.237	0.526	0.684	0.455	0.605	0.632	0.421	0.526	0.237	0.579	0.868	0.553	0.421	0.447	0.263	0.500	0.263	0.553	0.632	0.440	0.658	0.632	0.342	0.579	0.316	0.605	0.737	0.553
5	0.053	0.053	0.026	0.000	0.105	0.132	0.026	0.056	0.316	0.447	0.211	0.132	0.158	0.211	0.395	0.267	0.184	0.026	0.000	0.000	0.105	0.079	0.026	0.060	0.500	0.500	0.184	0.132	0.158	0.079	0.342	0.271
6	0.079	0.237	0.026	0.263	0.079	0.079	0.079		0.421	0.553	0.105	0.316	0.263	0.211	0.289	0.308	0.053	0.237	0.105	0.263	0.079	0.079	0.132	0.135	0.500	0.553	0.132	0.395	0.368	0.158	0.395	0.357
7	0.132	0.132	0.026	0.000	0.105	0.026	0.158	0.083	0.263	0.395	0.211	0.105	0.132	0.105	0.316	0.218	0.105	0.105	0.026	0.053	0.184	0.026	0.079	0.083	0.395	0.474	0.158	0.316	0.237	0.158	0.263	0.286
8	0.158	0.026	0.000	0.053	0.184	0.026	0.132	0.083	0.500	0.342	0.132	0.105	0.316	0.237	0.342	0.282	0.184	0.053	0.000	0.026	0.132	0.053	0.105	0.079	0.605	0.368	0.105	0.237	0.289	0.237	0.447	0.327
9	0.289	0.158	0.053	0.211	0.447	0.237	0.211	0.229	0.368	0.316	0.105	0.263	0.553	0.342	0.553	0.357	0.447	0.211	0.079	0.237	0.474	0.211	0.158	0.260	0.500	0.289	0.105	0.316	0.632	0.368	0.474	0.383
10	0.079	0.000	0.000	0.026	0.026	0.026	0.184		0.526	0.368	0.158	0.263	0.211	0.211	0.500	0.320	0.079	0.000	0.000	0.053	0.000	0.026	0.158	0.045	0.553	0.500	0.184	0.474	0.368	0.105	0.500	0.383
11	0.368	0.158	0.053	0.053	0.237	0.000	0.105	0.139	0.500	0.553	0.211	0.105	0.237	0.105	0.368	0.297	0.342	0.184	0.105	0.105	0.184	0.053	0.132	0.158	0.553	0.632	0.237	0.211	0.289	0.105	0.474	0.357
12	0.237	0.026	0.053	0.184	0.053	0.000	0.132	0.098	0.474	0.316	0.237	0.237	0.342	0.053	0.526	0.312	0.184	0.053	0.026	0.211	0.053	0.053	0.053		0.368	0.395	0.211	0.342	0.368	0.079	0.526	0.327
13	0.316	0.158	0.000	0.026	0.158	0.026	0.105	0.113	0.474	0.237	0.184	0.132	0.316	0.158	0.395	0.271	0.289	0.132	0.000	0.079	0.132	0.053	0.000	0.098	0.447	0.237	0.132	0.184	0.316	0.211	0.316	0.263
14	0.605	0.658	0.447	0.579	0.500	0.711	0.342	0.549	0.711	0.684	0.474	0.605	0.553	0.789	0.421	0.605	0.500	0.684	0.526	0.605	0.526	0.658	0.368	0.552	0.632	0.684	0.579	0.605	0.632	0.763	0.447	0.620
15	0.105	0.053	0.158	0.079	0.053	0.158	0.158	0.109	0.105	0.132	0.184	0.105	0.132	0.289	0.184	0.162	0.053	0.132	0.105	0.132	0.105	0.053	0.105	0.098	0.053	0.158	0.132	0.184	0.184	0.184	0.158	
16	0.158	0.026	0.105	0.053	0.079	0.053	0.079	0.079	0.342	0.158	0.184	0.132	0.132	0.211	0.158		0.132	0.026	0.079	0.026	0.132	0.053	0.026		0.368	0.079	0.184	0.079	0.158	0.184	0.105	0.165
17	0.395	0.053	0.053	0.237	0.237	0.000	0.184	0.166	0.500	0.421	0.211	0.316	0.395	0.105	0.553	0.357	0.500	0.079	0.079	0.289	0.289	0.026	0.158	0.203	0.526	0.474	0.184	0.395	0.342	0.132	0.526	0.368
18	**0.579**	**0.711**	**0.342**	**0.684**	**0.553**	**0.763**	**0.711**	0.620	0.711	0.737	0.395	0.711	0.579	0.763	0.868	0.681	**0.553**	**0.711**	**0.342**	**0.474**	**0.553**	**0.711**	**0.684**	0.575	0.658	0.737	0.395	0.553	0.605	0.763	0.816	0.647
19	0.474	0.605	0.237	0.684	0.263	0.605	0.632	0.500	0.737	0.605	0.263	0.737	0.579	0.711	0.763	0.628	0.474	0.605	0.211	0.526	0.316	0.579	0.579	0.470	0.684	0.632	0.263	0.737	0.632	0.763	0.711	
20	0.605	0.684	0.263	0.368	0.447	0.395	0.553	0.474	0.711	0.789	0.342	0.553	0.474	0.474	0.658	0.572	0.526	0.684	0.316	0.368	0.526	0.395	0.526	0.477	0.658	0.737	0.395	0.526	0.632	0.421	0.658	0.575
21	0.342	0.526	0.263	0.158	0.395	0.711	0.553	0.421	0.421	0.605	0.368	0.289	0.421	0.737	0.658	0.500	0.342	0.605	0.289	0.237	0.447	0.684	0.395	0.428	0.421	0.605	0.395	0.395	0.500	0.763	0.447	0.504
22	0.079	0.000	0.053	0.053	0.026	0.000	0.000	0.030	0.184	0.263	0.237	0.184	0.105	0.263	0.158	0.199	0.158	0.026	0.053	0.053	0.053	0.026	0.026		0.237	0.132	0.158	0.211	0.132	0.237	0.132	
Total								0.275								0.416								0.302								0.434
PreMod	**0.579**	**0.711**	0.447	0.579	**0.553**	**0.763**	**0.711**	0.620	0.711	0.737	0.474	0.605	0.579	0.763	0.868	0.677	**0.553**	**0.711**	0.526	0.605	**0.553**	**0.711**	**0.684**	0.620	0.658	0.737	0.579	0.605	0.605	0.763	0.816	0.680

Note: *The number "1" and "2" donate that the predicting helical regions of GPCRs by the scoring matrices of the training sets are "partial consistent with" and "identical with" the actual TM regions of GPCRs, respectively.

Table 6. The t-test results based on the scores and the validity of GPCRs in test sets by training set 18, 14 and 3.

Training Set	S^L_{D18}													S^L_{A14}												S^L_{A3}										
Validity	2					2 + 1					0			2					2 + 1					0			2					2 + 1					0		
t-test	n_1	$\bar{x}_1$	S_1	t	P	n_1	$\bar{x}_1$	S_1	t	P	n_2	$\bar{x}_2$	S_2	n_1	$\bar{x}_1$	S_1	t	P	n_1	$\bar{x}_1$	S_1	t	P	n_2	$\bar{x}_2$	S_2	n_1	$\bar{x}_1$	S_1	t	P	n_1	$\bar{x}_1$	S_1	t	P	n_2	$\bar{x}_2$	S_2
Test set 23																																						
TM1	22	0.421	0.029	2.378	<0.05	27	0.421	0.029	2.417	<0.05	11	0.398	0.019	23	0.479	0.045	3.402	<0.01	27	0.477	0.045	3.288	<0.01	11	0.430	0.022	19	0.503	0.052	4.043	<0.001	28	0.494	0.049	3.777	<0.001	10	0.433	0.022
TM2	27	0.465	0.027	6.565	<0.001	28	0.461	0.032	5.409	<0.001	10	0.400	0.026	25	0.548	0.040	6.959	<0.001	26	0.543	0.046	5.851	<0.001	12	0.459	0.027	21	0.519	0.054	5.705	<0.001	23	0.517	0.054	5.590	<0.001	15	0.435	0.021
TM3	13	0.410	0.020	3.314	<0.01	15	0.414	0.022	3.911	<0.001	23	0.387	0.020	17	0.454	0.021	6.817	<0.001	18	0.455	0.020	7.209	<0.001	20	0.403	0.024	14	0.508	0.058	5.528	<0.001	16	0.505	0.059	5.266	<0.001	22	0.434	0.020
TM4	26	0.402	0.024	1.238	>0.05	27	0.401	0.024	1.119	>0.05	11	0.392	0.018	22	0.447	0.023	4.137	<0.001	23	0.446	0.024	3.953	<0.001	15	0.414	0.025	21	0.475	0.033	2.369	<0.001	23	0.478	0.039	2.376	<0.001	15	0.451	0.025
TM5	21	0.438	0.030	3.452	<0.01	22	0.437	0.030	3.356	<0.01	16	0.408	0.020	19	0.464	0.025	5.631	<0.001	21	0.464	0.026	5.636	<0.001	17	0.417	0.025	21	0.478	0.037	3.816	<0.001	24	0.477	0.040	3.566	<0.001	14	0.434	0.027
TM6	29	0.464	0.036	3.231	<0.01	29	0.464	0.036	3.231	<0.01	9	0.423	0.021	27	0.509	0.049	3.300	<0.01	30	0.507	0.047	3.333	<0.01	8	0.448	0.032	28	0.500	0.049	3.522	<0.001	31	0.496	0.053	3.088	<0.001	7	0.432	0.026
TM7	27	0.481	0.046	3.594	<0.01	33	0.481	0.048	3.451	<0.01	5	0.406	0.005	13	0.443	0.027	3.282	<0.01	16	0.445	0.035	3.283	<0.01	22	0.412	0.027	24	0.550	0.070	4.212	<0.001	29	0.552	0.071	4.259	<0.001	9	0.448	0.028
Test set 24																																						
TM1	21	0.426	0.029	3.787	<0.001	25	0.427	0.029	3.958	<0.001	13	0.392	0.018	19	0.488	0.036	5.286	<0.001	23	0.482	0.038	4.604	<0.001	15	0.432	0.022	18	0.500	0.056	3.960	<0.001	26	0.500	0.050	4.453	<0.001	12	0.433	0.020
TM2	27	0.465	0.027	6.116	<0.001	28	0.463	0.027	5.938	<0.001	10	0.405	0.025	26	0.540	0.042	5.623	<0.001	26	0.540	0.042	5.623	<0.001	12	0.464	0.030	23	0.508	0.053	4.553	<0.001	25	0.507	0.053	4.494	<0.001	13	0.439	0.016
TM3	13	0.407	0.021	3.269	<0.01	15	0.408	0.025	3.236	<0.01	23	0.386	0.017	20	0.448	0.021	6.524	<0.001	22	0.449	0.021	6.800	<0.001	16	0.403	0.020	12	0.495	0.052	4.630	<0.001	15	0.488	0.050	4.275	<0.001	23	0.438	0.021
TM4	18	0.411	0.019	3.602	<0.01	21	0.409	0.020	3.274	<0.01	17	0.389	0.017	23	0.450	0.029	4.883	<0.001	23	0.450	0.029	4.883	<0.001	15	0.409	0.018	18	0.483	0.035	3.247	<0.001	21	0.483	0.041	2.966	<0.001	17	0.449	0.026
TM5	21	0.441	0.026	4.008	<0.001	23	0.439	0.026	3.771	<0.001	15	0.411	0.015	20	0.466	0.028	4.181	<0.001	24	0.462	0.031	3.563	<0.01	14	0.428	0.023	20	0.469	0.038	2.310	<0.05	26	0.465	0.041	1.921	>0.05	12	0.440	0.027
TM6	27	0.474	0.037	5.044	<0.001	29	0.470	0.038	4.616	<0.001	9	0.410	0.013	25	0.523	0.050	5.548	<0.001	29	0.518	0.048	5.476	<0.001	9	0.429	0.010	25	0.509	0.049	5.215	<0.001	30	0.507	0.050	5.010	<0.001	8	0.417	0.011
TM7	26	0.478	0.045	4.041	<0.001	31	0.476	0.045	3.944	<0.001	7	0.407	0.019	14	0.444	0.030	4.282	<0.001	17	0.441	0.031	3.920	<0.001	21	0.410	0.017	24	0.535	0.072	3.525	<0.001	27	0.537	0.069	3.775	<0.001	11	0.456	0.024

Note: "n", "$\bar{x}$", and "S" donate the sample number, mean, and standard deviation (S.D.) of the scores of GPCRs in the test set (T23 and T24) by the scoring matrices of training set 18, 14 and 3. $\bar{x} = \sum x/n$; $S = \sqrt{\left(n \times \sum x^2 - \left(\sum x\right)^2\right) \Big/ \left(n(n-1)\right)}$; $t = \left|\bar{x}_1 - \bar{x}_2\right| \Big/ \sqrt{\left\{\left[(n_1-1)S_1^2 + (n_2-1)S_2^2\right] \Big/ (n_1+n_2-2)\right\}(1/n_1 + 1/n_2)}$.

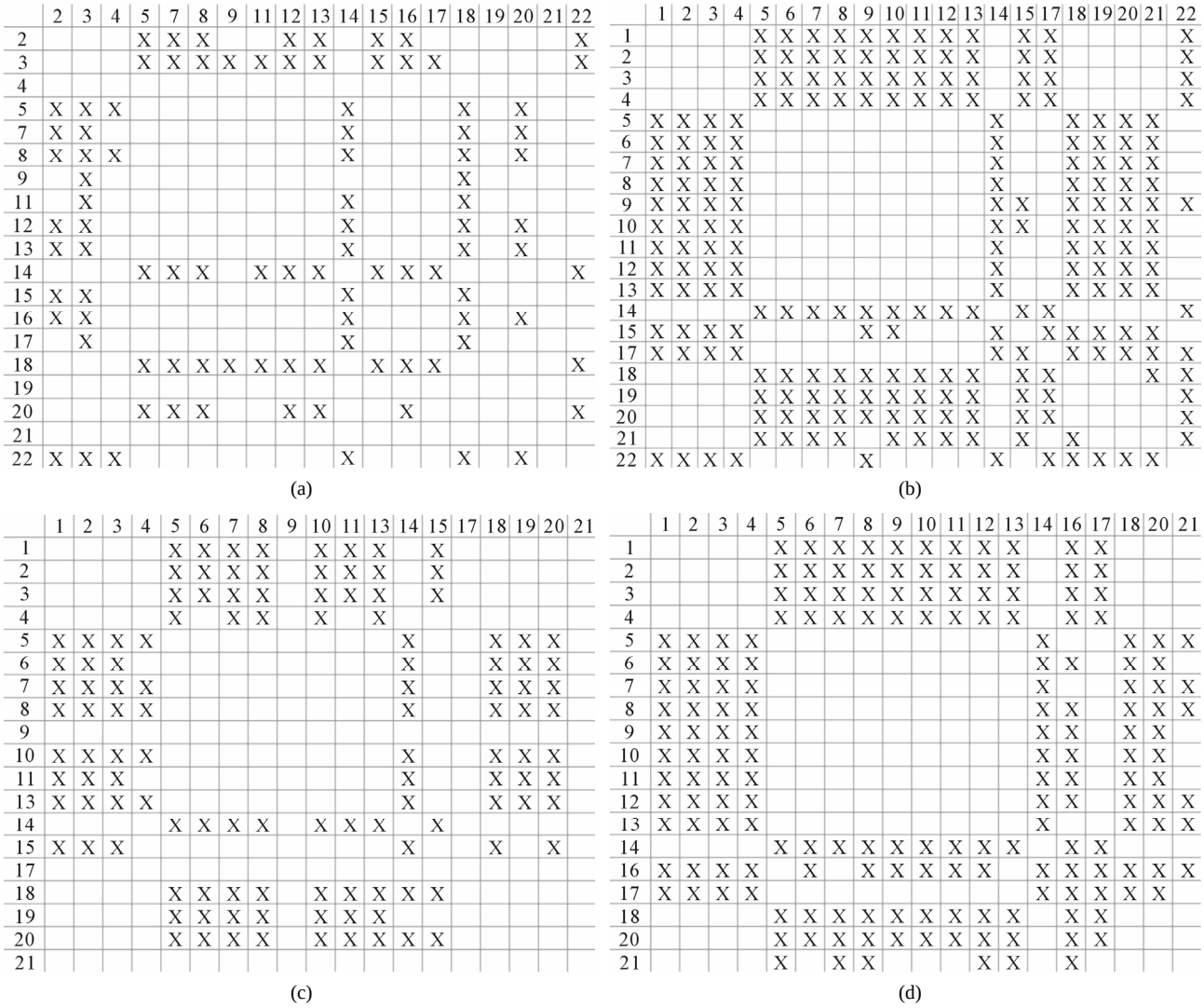

Figure 4. The results of multiple comparisons, where "X" shows there are significant differences (at the significant level of 0.05) between the two examinations indicated by corresponding column and line. (a) S_{23}^{T}-"2"; (b) S_{23}^{T}-"1 + 2"; (c) S_{24}^{T}-"2"; and (d) S_{24}^{T}-"1 + 2".

gram of the accuracy of the coding sequences of GPCRs' TMs in test sets (S_{23}^{T} and S_{24}^{T}) by the scoring matrices of the different training datasets (S_{D18}^{L}, S_{A14}^{L}, and S_{A3}^{L}). At the complete consistent "2" state, the accuracy of TM1, TM2, TM5, TM6 and TM7 of GPCRs in S_{24}^{T} by S_{D18}^{L} are higher than others, while that of TM3 and TM4 by S_{A14}^{L} are higher than others; the accuracy of TM2, TM4, TM5, TM6 and TM7 of GPCRs in S_{23}^{T} by S_{D18}^{L} are higher than others while that of TM1 and TM3 by S_{A14}^{L} are higher than others. At the crossover "2 + 1" state, the accuracy of TM2 and TM7 of GPCRs in S_{23}^{T} and S_{24}^{T} by S_{D18}^{L} are higher; that of TM1, TM5, and TM6 of GPCRs in S_{23}^{T} and S_{24}^{T} by S_{A3}^{L} are higher; and that of TM3 of GPCRs in S_{23}^{T} and S_{24}^{T} by S_{A14}^{L} is higher, but that of TM4 of GPCRs in S_{23}^{T} by S_{D18}^{L} and in S_{24}^{T} by A14 are higher. Moreover, the accuracy of TM1-TM7 of GPCRs in S_{24}^{T} by S_{D18}^{L}, S_{A14}^{L} and S_{A3}^{L} are higher than that of GPCRs in S_{23}^{T}, respectively.

Considering the mean scores, the accuracy, and t-test values mentioned above, the scoring matrices of seven TMs of GPCRs is composed of the scoring matrices of TM1, TM2, TM5, TM6 and TM7 from S_{D18}^{L} and those of TM3 and TM4 from S_{A14}^{L}. So we constructed a prediction model (PreMod) based on the combined scoring matrices for prediction of potential GPCR proteins and further helpful to discovery of potential drug targets belonging to GPCR families. Comparison the prediction values of two test sets using PreMod (**Table 7**) with those of other models (S_{D18}^{L}, S_{A14}^{L} and S_{A3}^{L}) (**Table 3**) reveals that the prediction seven TMs' accuracy of the former is more than the latter's whereas the hit rate of the former (94.74% and 97.37%) is less than that of S_{A3}^{L} (100%) but bigger than that of S_{D18}^{L} and S_{A14}^{L} (**Table 8**). This is the reason that we choose PreMod to predict

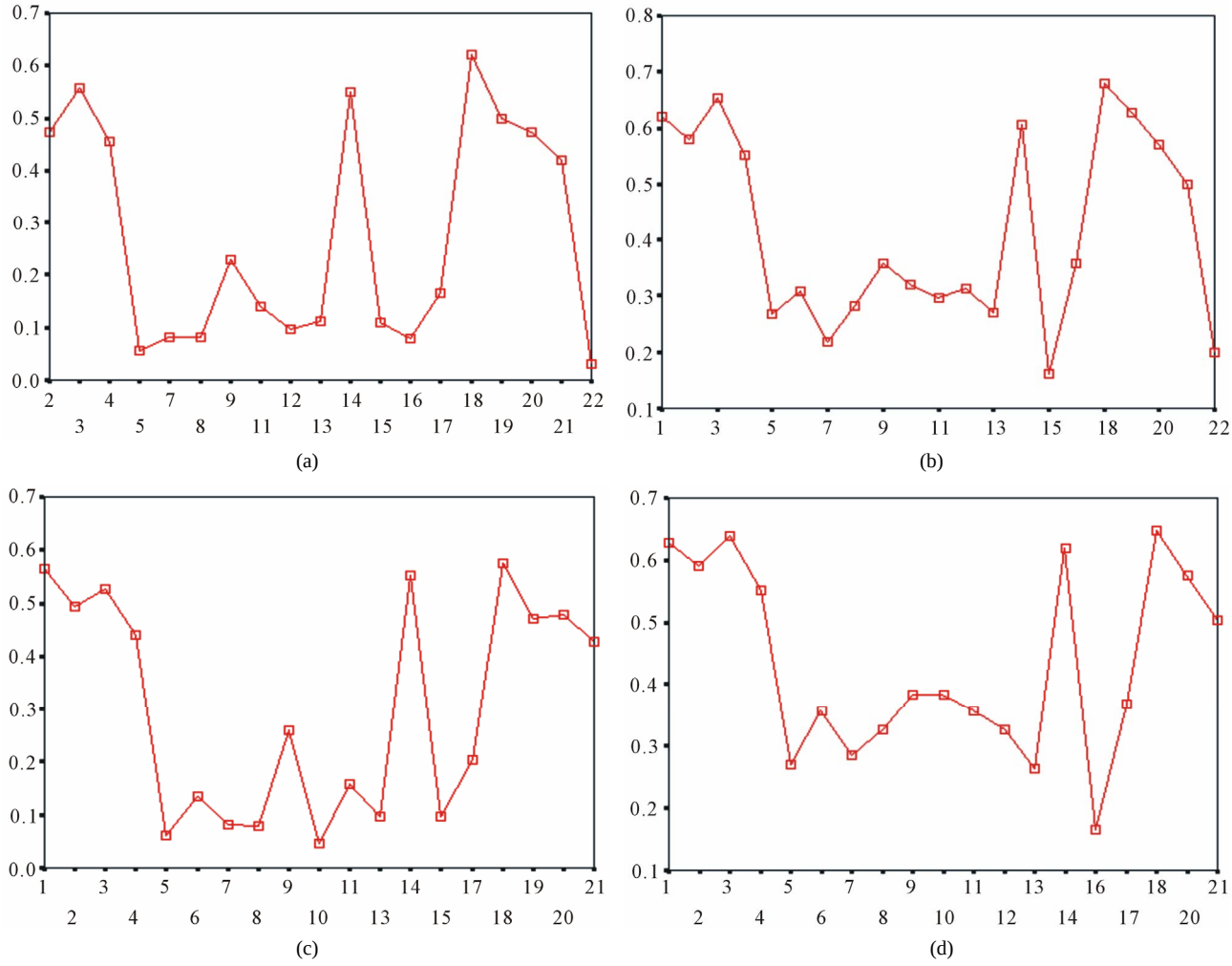

Figure 5. The plot of the averages of validity of prediction each training set (examinations of each category) by two test sets (S_{23}^{T} and S_{24}^{T}), where X-coordinate and Y-coordinate axes means examinations and the averages of validity of prediction, respectively. (a) S_{23}^{T}-"2"; (b) S_{23}^{T}-"1 + 2"; (c) S_{24}^{T}-"2"; and (d) S_{24}^{T}-"1 + 2".

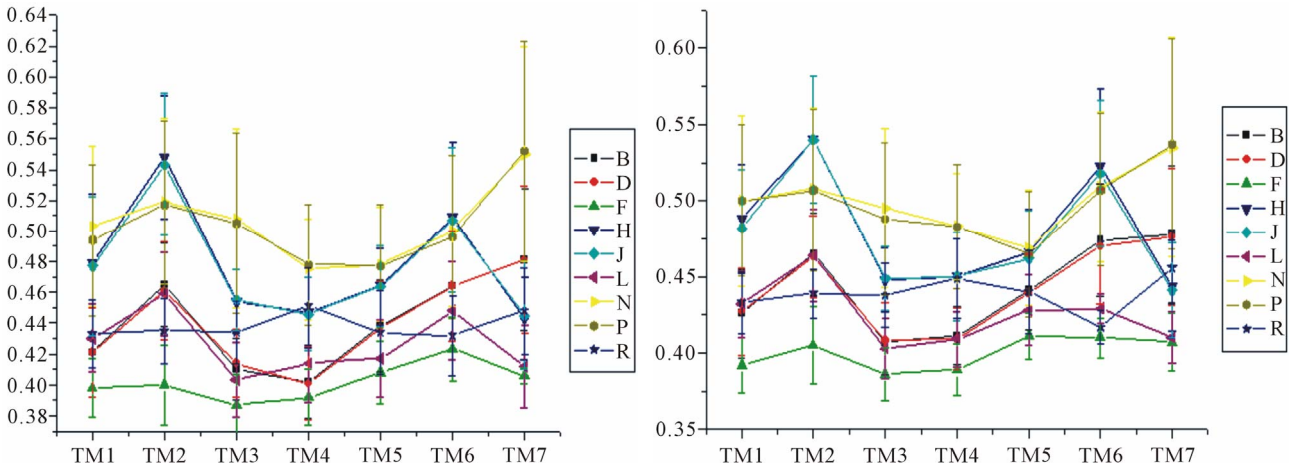

Figure 6. The mean score of the coding sequences of GPCRs' 7 TMs in test sets (T23 and T24) by the scoring matrices of the different training datasets (D18, A14, and A3). Upper: T23; Lower: T24. Here, B, D, and F lines express D18; H, J, and L, A14; and N, P, and R, A3. B, H, and N means "2"; D, J, and P, "2 + 1"; and F, L, and R, "0".

Table 7. The score and the validity of the coding sequences of GPCRs' each trans-membrane segment in test sets by the scoring matrices of PreMod.

GPCRs	Score							Validity							Accumulation				GPCRs	Score							Validity							Accumulation			
Test set 23	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	"2"	"1"	"2 + 1"	"0"	Test set 24	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	"2"	"1"	"2 + 1"	"0"
Q5VTM0_HUMAN	0.473	0.494	0.427	0.423	0.424	0.437	0.458	1	2	2	2	2	2	2	6	1	7	0	Q5VTM0_HUMAN	0.473	0.494	0.427	0.423	0.424	0.437	0.458	1	2	2	2	2	2	2	6	1	7	0
NPY5R_HUMAN	0.424	0.464	0.371	0.395	0.384	0.426	0.434	2	2	0	2	0	2	2	5	0	5	2	BRS3_HUMAN	0.398	0.451	0.372	0.462	0.441	0.498	0.415	1	2	0	2	2	2	2	5	1	6	1
BRS3_HUMAN	0.398	0.451	0.372	0.462	0.441	0.498	0.415	1	2	0	2	2	2	2	5	1	6	1	GP173_HUMAN	0.391	0.406	0.402	0.455	0.411	0.454	0.454	2	2	0	2	0	2	2	5	0	5	2
LGR8_HUMAN	0.392	0.418	0.409	0.450	0.405	0.425	0.454	2	0	2	0	0	2	1	3	1	4	3	EDG8_HUMAN	0.425	0.444	0.451	0.460	0.419	0.468	0.514	0	2	0	2	2	2	2	5	0	5	2
GP173_HUMAN	0.391	0.406	0.402	0.455	0.411	0.454	0.454	2	2	0	2	0	2	2	5	0	5	2	ADA2C_HUMAN	0.458	0.464	0.435	0.486	0.432	0.547	0.539	2	2	0	2	2	2	2	6	0	6	1
OPSB_HUMAN	0.409	0.461	0.434	0.429	0.484	0.439	0.453	2	2	0	0	2	2	2	5	0	5	2	5HT1E_HUMAN	0.398	0.466	0.470	0.469	0.428	0.480	0.502	0	2	2	2	0	2	2	5	0	5	2
EDG8_HUMAN	0.425	0.444	0.451	0.460	0.419	0.468	0.514	0	2	0	2	2	2	2	5	0	5	2	TAAR2_HUMAN	0.363	0.461	0.403	0.399	0.404	0.440	0.446	0	2	0	0	0	2	2	3	0	3	4
PF2R_HUMAN	0.386	0.373	0.367	0.393	0.389	0.414	0.383	2	0	0	0	0	0	2	2	0	2	5	ACM5_HUMAN	0.446	0.449	0.438	0.434	0.432	0.460	0.498	2	2	2	2	2	2	2	7	0	7	0
ADA2C_HUMAN	0.458	0.464	0.435	0.486	0.432	0.547	0.539	2	2	0	2	2	2	2	6	0	6	1	CXCR6_HUMAN	0.436	0.490	0.477	0.491	0.468	0.443	0.473	2	2	2	2	2	2	2	7	0	7	0
ADRB3_HUMAN	0.407	0.442	0.440	0.457	0.473	0.518	0.505	2	2	2	2	2	2	2	7	0	7	0	OPRX_HUMAN	0.448	0.490	0.475	0.465	0.496	0.484	0.571	2	2	2	2	2	2	2	7	0	7	0
5HT1E_HUMAN	0.398	0.466	0.470	0.469	0.428	0.480	0.502	0	2	2	2	0	2	2	5	0	5	2	C5ARL_HUMAN	0.475	0.476	0.473	0.452	0.423	0.516	0.486	2	2	2	2	2	2	2	7	0	7	0
5HT6R_HUMAN	0.398	0.420	0.492	0.439	0.424	0.478	0.536	0	2	2	2	2	2	1	5	1	6	1	P2Y12_HUMAN	0.422	0.415	0.388	0.429	0.413	0.441	0.461	1	0	0	2	0	2	2	3	1	4	3
TAAR2_HUMAN	0.363	0.461	0.403	0.399	0.404	0.440	0.446	0	2	0	0	0	2	2	3	0	3	4	PAR4_HUMAN	0.439	0.481	0.466	0.427	0.437	0.421	0.529	2	2	2	0	2	2	2	6	0	6	1
AA3R_HUMAN	0.400	0.447	0.446	0.416	0.421	0.434	0.508	2	2	2	2	0	2	2	6	0	6	1	EMR1_HUMAN	0.392	0.401	0.414	0.421	0.417	0.430	0.444	0	0	0	0	0	0	1	0	1	1	6
ACM5_HUMAN	0.446	0.449	0.438	0.434	0.432	0.460	0.498	2	2	2	2	2	2	2	7	0	7	0	MGR8_HUMAN	0.386	0.382	0.405	0.388	0.397	0.421	0.399	2	0	0	0	0	0	0	1	0	1	6
HRH4_HUMAN	0.420	0.468	0.425	0.452	0.402	0.450	0.471	0	2	0	2	2	2	2	5	0	5	2	FZD6_HUMAN	0.387	0.368	0.381	0.390	0.394	0.402	0.408	0	0	0	0	0	0	0	0	0	0	7
CXCR6_HUMAN	0.436	0.490	0.477	0.491	0.468	0.443	0.473	2	2	2	2	2	2	2	7	0	7	0	NPBW2_HUMAN	0.410	0.500	0.456	0.427	0.444	0.508	0.523	2	2	2	2	2	2	1	6	1	7	0
SSR5_HUMAN	0.422	0.495	0.459	0.468	0.463	0.484	0.512	2	2	2	2	2	2	2	7	0	7	0	VN1R2_HUMAN	0.374	0.434	0.402	0.406	0.392	0.401	0.406	0	0	0	0	0	0	0	0	0	0	7
OPRX_HUMAN	0.448	0.490	0.475	0.465	0.496	0.484	0.571	2	2	2	2	2	2	2	7	0	7	0	OR2C1_HUMAN	0.398	0.412	0.396	0.416	0.405	0.413	0.454	0	2	0	0	2	2	2	4	0	4	3
FPR1_HUMAN	0.457	0.470	0.439	0.447	0.428	0.456	0.560	2	2	2	2	2	2	2	7	0	7	0	AGTR1_BOVIN	0.453	0.418	0.411	0.388	0.435	0.474	0.479	2	2	2	2	2	2	1	6	1	7	0
C5ARL_HUMAN	0.475	0.476	0.473	0.452	0.423	0.516	0.486	2	2	2	2	2	2	2	7	0	7	0	Q3T181_BOVIN	0.394	0.475	0.456	0.502	0.452	0.517	0.510	0	2	2	2	0	2	2	5	0	5	2
GALR3_HUMAN	0.461	0.502	0.454	0.453	0.400	0.466	0.538	2	2	2	2	2	0	1	5	1	6	1	FSHR_BOVIN	0.415	0.467	0.425	0.424	0.415	0.426	0.480	2	2	2	2	0	0	2	5	0	5	2
P2Y12_HUMAN	0.422	0.415	0.388	0.429	0.413	0.441	0.461	1	0	0	2	0	2	2	3	1	4	3	GNRHR_CANFA	0.374	0.437	0.461	0.409	0.451	0.429	0.434	0	2	2	0	2	1	2	4	1	5	2
P2RY4_HUMAN	0.410	0.442	0.443	0.422	0.397	0.494	0.508	0	2	2	0	2	2	2	5	0	5	2	EDNRA_CANFA	0.392	0.460	0.428	0.469	0.404	0.475	0.401	2	2	2	2	0	2	2	6	0	6	1
PAR4_HUMAN	0.439	0.481	0.466	0.427	0.437	0.421	0.529	2	2	2	0	2	2	2	6	0	6	1	5HT1A_CANFA	0.400	0.465	0.452	0.440	0.454	0.544	0.552	2	2	2	2	2	2	2	7	0	7	0
MRGRD_HUMAN	0.410	0.509	0.410	0.426	0.395	0.456	0.462	1	2	0	0	0	2	2	3	1	4	3	OPS5_DROME	0.434	0.485	0.385	0.413	0.426	0.417	0.390	2	2	0	0	0	2	0	3	0	3	4
EMR1_HUMAN	0.392	0.401	0.414	0.421	0.417	0.430	0.444	0	0	0	0	0	0	1	0	1	1	6	GR33A_DROME	0.410	0.396	0.392	0.400	0.403	0.414	0.445	0	0	0	0	1	0	0	1	0	1	6
GLR_HUMAN	0.418	0.444	0.421	0.431	0.443	0.438	0.472	0	0	0	2	2	0	2	3	0	3	4	MTH4_DROME	0.369	0.407	0.424	0.398	0.400	0.394	0.397	0	0	2	2	0	0	1	2	1	3	4
MGR8_HUMAN	0.386	0.382	0.405	0.388	0.397	0.421	0.399	2	0	0	0	0	0	0	1	0	1	6	OLF6_CHICK	0.415	0.404	0.400	0.380	0.400	0.398	0.444	0	0	0	0	2	0	2	2	0	2	5
GPC5B_HUMAN	0.397	0.413	0.424	0.454	0.463	0.429	0.397	0	2	0	0	0	0	1	1	1	2	5	FZD7_CHICK	0.392	0.453	0.429	0.430	0.413	0.410	0.414	0	0	2	0	0	1	0	1	1	2	5
FZD6_HUMAN	0.387	0.368	0.381	0.390	0.394	0.402	0.408	0	0	0	0	0	0	0	0	0	0	7	P2RY1_CHICK	0.453	0.514	0.468	0.450	0.448	0.493	0.551	2	2	2	2	2	2	2	7	0	7	0
CML2_HUMAN	0.437	0.494	0.460	0.433	0.484	0.517	0.516	2	2	2	2	2	2	2	7	0	7	0	TAAR6_MOUSE	0.441	0.442	0.432	0.451	0.405	0.477	0.477	1	2	2	2	2	2	1	5	2	7	0
NPBW2_HUMAN	0.410	0.500	0.456	0.427	0.444	0.508	0.523	2	2	2	2	2	2	1	6	1	7	0	NMBR_MOUSE	0.461	0.460	0.427	0.488	0.471	0.507	0.415	2	2	2	2	2	2	2	7	0	7	0
P2Y10_HUMAN	0.376	0.448	0.461	0.442	0.407	0.460	0.460	2	2	1	0	1	2	2	4	2	6	1	FFAR2_MOUSE	0.423	0.478	0.438	0.430	0.431	0.449	0.452	2	2	1	2	2	2	2	6	1	7	0
GP158_HUMAN	0.389	0.390	0.390	0.403	0.393	0.406	0.412	2	0	0	1	0	0	0	1	1	2	5	MLO10_ARATH	0.384	0.385	0.402	0.446	0.405	0.406	0.388	2	0	0	0	0	0	0	1	0	1	6
T2R43_HUMAN	0.395	0.371	0.376	0.373	0.393	0.402	0.388	1	1	0	0	2	2	2	3	2	5	2	5HT2A_MACMU	0.389	0.453	0.464	0.395	0.462	0.442	0.454	2	2	2	0	2	2	2	6	0	6	1
VN1R2_HUMAN	0.374	0.434	0.402	0.406	0.392	0.401	0.406	0	0	0	0	0	0	0	0	0	0	0	OPSD_MESBI	0.402	0.505	0.421	0.468	0.492	0.506	0.457	2	2	0	2	2	2	2	6	0	6	1
O52A1_HUMAN	0.408	0.378	0.388	0.387	0.420	0.409	0.405	2	0	0	0	0	2	0	2	0	2	5	V1AR_MICOH	0.449	0.432	0.470	0.417	0.426	0.490	0.490	2	1	1	0	1	2	2	3	3	6	1
"2"	0.421 ± 0.029	0.465 ± 0.027	0.454 ± 0.021	0.447 ± 0.023	0.438 ± 0.030	0.464 ± 0.036	0.481 ± 0.046	0.579	0.711	0.447	0.579	0.553	0.763	0.711	0.211				"2"	0.426 ± 0.029	0.465 ± 0.027	0.448 ± 0.021	0.450 ± 0.029	0.441 ± 0.026	0.474 ± 0.037	0.478 ± 0.045	0.553	0.711	0.526	0.605	0.553	0.711	0.684	0.184			
"2 +1"	0.421 ± 0.029	0.461 ± 0.032	0.455 ± 0.020	0.446 ± 0.024	0.437 ± 0.030	0.464 ± 0.036	0.481 ± 0.048	0.711	0.737	0.474	0.605	0.579	0.763	0.868			0.263		"2 + 1"	0.427 ± 0.029	0.463 ± 0.027	0.449 ± 0.021	0.450 ± 0.029	0.439 ± 0.026	0.470 ± 0.038	0.476 ± 0.045	0.658	0.737	0.579	0.605	0.605	0.763	0.816			0.316	
"0"																	0.026		"0"																0.053

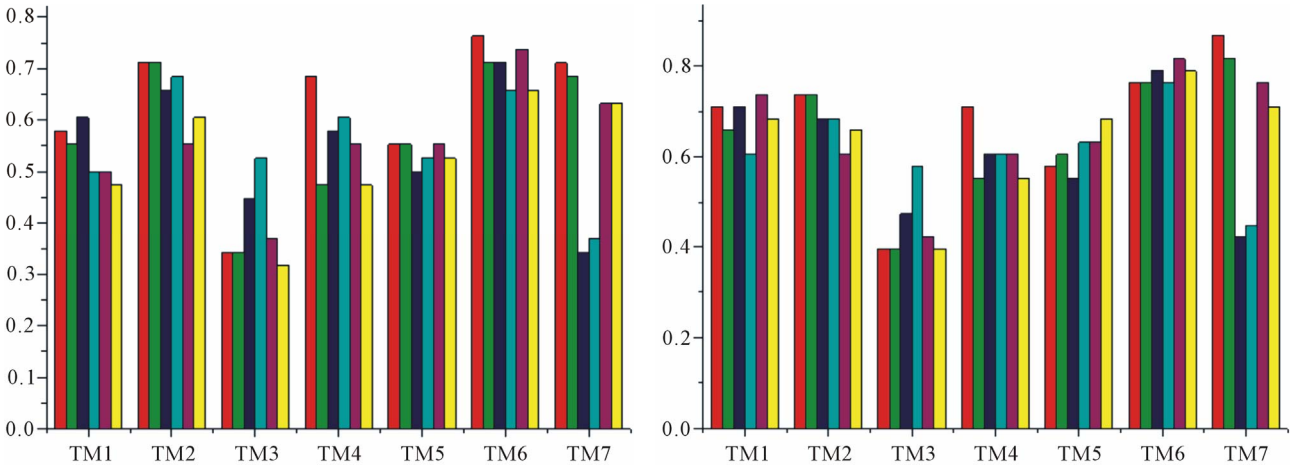

Figure 7. The accuracy of the coding sequences of GPCRs' TMs in test sets by the scoring matrices of the different training datasets. Left: "2"; Right: "2 + 1". Here, red and green poles display D18; blue and cyan, A14; and magenta and yellow, A3. Red, Blue, and magenta mean T23 while the rest do T24.

Table 8. The validity and hit rates of prediction all seven transmembrane regions of GPCRs in test sets by the scoring matrices of training set 18, 14, 3 and PreMod.

Model	D18				A14				A3				PreMod			
	Accuracy	Validity	hits	miss	Accuracy	Validity	hit	miss	Accuracy	Validity	hit	miss	Accuracy	Validity	hit	miss
	"2"	"2+1"	"2 "or "1"	"0"	"2"	"2 + 1"	"2 "or "1"	"0"	"2"	"2+1"	"2 "or "1"	"0"	"2"	"2+1"	"2 "or "1"	"0"
	Times															
Test ser 23	6	8	36	2	5	8	34	4	3	7	38	0	8	10	37	1
Test set 24	4	7	36	2	4	7	34	4	1	6	38	0	7	12	36	2
	Percent (%)															
Test set 23	15.79	21.05	94.74	5.26	13.16	21.05	89.47	10.53	7.89	18.42	100	0	21.05	26.32	97.37	2.63
Test set 24	10.53	18.42	94.74	5.26	10.53	18.42	89.47	10.53	2.63	15.79	100	0	18.42	31.58	94.74	5.26

Note: Here, these numbers are all seven TMs.

some potential drug targets.

The test of the normality of scores of Test sets by the Shapiro-Wilk statistic reveals that the data of model 18 and 14 fit the normal distribution and all statistical testing was conducted at significance level 0.10 with all confidence intervals at confidence level 0.90 (**Table 9**). Here, the W values are between zero and one, and typical value is 0.10. The significative p-values are more than the default value 0.05 at the level of $\alpha = 0.10$. The batch means pass the Shapiro-Wilk test for multivariate normality. Particularly, the means of scores on the basis of model 18 are less than those of model 14 while "2 + 1" combined with Test set 24 model is superior to "2" binding to Test set 23 model. Take 90% confidence interval lower limit of "2 + 1" combined with Test set 24 model based on model 18 as the threshold of TM1-7 of GPCRs for prediction of chromosome 19 (Validation Set).

Table 10 displays the score and the validity of the coding sequences of each trans-membrane segment of 22 GPCRs located in sense chain of chromosome 19 by the scoring matrices of Group 18 (model 18). Chromosome 19 is composed of four configs: Config 1, 2, 3, and 4, containing 3, 0, 11, and 7 GPCRs, respectively. The prediction results of these GPCRs show that there are 19, 8, 9, 5, 6, 6, and 11 positive data of TM1, TM2, TM3, TM4, TM5, TM6, and TM7, respectively. Plot of predicted and actual values of GPCRs in four configs of chromosome 19 shows that four TMs (*i.e.* TM1, TM2, TM6, and TM7) have higher prediction accuracy while other three TMs (such as TM3, TM4, and TM5) possess lower positive results (**Figure 8**), especially TM1 and TM7 with positive rates of 19/74 and 11/25, respectively. However, the "hits" rate is up to 20/22 if only anyone TM fits to the actual TMs of GPCRs.

Due to the size of our data sets, pairwise computation

Table 9. Normality analysis of the scores of 38 GPCRs' 7 TMs in Test set 24 by the scoring matrices of PreMod.

GPCRs	Test set 23								Test set 24							
	Score			90% Confidence Interval for Mean			Shapiro-Wilk		Score			90% Confidence Interval for Mean			Shapiro-Wilk	
	Mean	S.D.	D.E.	Lower	Upper	df	StatisticW value	Sig. P-value	Mean	S.D.	D.E.	Lower	Upper	df	StatisticW value	Sig. P-value
Group 18																
"2"																
TM1	0.421	0.029	0.006	0.410	0.431	22	0.949	0.296	0.426	0.029	0.006	0.415	0.437	21	0.924	0.107
TM2	0.465	0.027	0.005	0.474	0.556	27	0.961	0.382	0.465	0.027	0.005	0.456	0.473	27	0.976	0.771
TM3	0.410	0.020	0.005	0.400	0.420	13	0.907	0.169	0.407	0.021	0.006	0.397	0.418	13	0.946	0.545
TM4	0.402	0.024	0.005	0.394	0.410	26	0.974	0.717	0.411	0.019	0.004	0.403	0.418	18	0.930	0.195
TM5	0.438	0.030	0.007	0.427	0.450	21	0.950	0.336	0.441	0.026	0.006	0.432	0.451	21	0.963	0.587
TM6	0.464	0.036	0.007	0.452	0.475	29	0.970	0.568	0.474	0.037	0.007	0.462	0.481	27	0.971	0.632
TM7	0.481	0.046	0.009	0.465	0.496	27	0.976	0.765	0.478	0.045	0.009	0.463	0.494	26	0.964	0.466
Mean of 7 TMs' mean	0.440	0.031	0.012	0.418	0.463	7	0.926	0.519	0.443	0.030	0.011	0.421	0.465	7	0.897	0.313
Mean of 7 TMs' S.D.	0.030	0.009	0.003	0.024	0.037	7	0.938	0.624	0.029	0.009	0.003	0.022	0.036	7	0.928	0.534
"2 + 1"																
TM1	0.421	0.029	0.006	0.411	0.430	27	0.944	0.150	0.427	0.029	0.006	0.417	0.437	25	0.937	0.125
TM2	0.461	0.032	0.006	0.451	0.472	28	0.942	0.126	0.463	0.027	0.005	0.455	0.472	28	0.982	0.905
TM3	0.414	0.022	0.006	0.405	0.424	15	0.940	0.380	0.408	0.025	0.006	0.396	0.419	15	0.963	0.738
TM4	0.401	0.024	0.005	0.393	0.409	27	0.966	0.494	0.409	0.020	0.004	0.401	0.416	21	0.935	0.175
TM5	0.437	0.030	0.006	0.426	0.448	22	0.947	0.277	0.439	0.026	0.005	0.430	0.449	23	0.955	0.375
TM6	0.464	0.036	0.007	0.452	0.475	29	0.970	0.568	0.470	0.038	0.007	0.458	0.482	29	0.967	0.478
TM7	0.481	0.048	0.008	0.467	0.495	33	0.974	0.589	0.476	0.045	0.008	0.462	0.490	31	0.974	0.622
Mean of 7 TMs' mean	0.440	0.030	0.011	0.418	0.462	7	0.948	0.711	0.442	0.028	0.011	0.421	0.463	7	0.895	0.300
Mean of 7 TMs' S.D.	0.032	0.009	0.003	0.025	0.038	7	0.917	0.445	0.030	0.009	0.003	0.024	0.036	7	0.904	0.354
Group 14																
"2"																
TM1	0.479	0.045	0.009	0.462	0.495	23	0.932	0.123	0.488	0.036	0.008	0.473	0.502	19	0.979	0.930
TM2	0.548	0.040	0.008	0.534	0.561	25	0.918	0.046	0.540	0.042	0.008	0.526	0.554	26	0.933	0.089
TM3	0.454	0.021	0.005	0.446	0.463	17	0.983	0.978	0.448	0.021	0.005	0.440	0.456	20	0.903	0.046
TM4	0.447	0.023	0.005	0.439	0.456	22	0.982	0.941	0.450	0.029	0.006	0.440	0.461	23	0.976	0.821
TM5	0.464	0.025	0.006	0.454	0.474	19	0.964	0.647	0.466	0.028	0.006	0.455	0.466	20	0.952	0.405
TM6	0.509	0.049	0.009	0.493	0.525	27	0.969	0.575	0.523	0.050	0.010	0.506	0.540	25	0.968	0.599
TM7	0.443	0.027	0.008	0.429	0.456	13	0.992	0.999	0.444	0.030	0.008	0.429	0.458	14	0.944	0.467
Mean of 7 TMs' mean	0.478	0.038	0.014	0.450	0.506	7	0.873	0.197	0.480	0.039	0.015	0.452	0.508	7	0.865	0.169
Mean of 7 TMs' S.D.	0.033	0.011	0.004	0.024	0.041	7	0.868	0.178	0.034	0.010	0.004	0.027	0.041	7	0.953	0.761

Continued

"2 + 1"																
TM1	0.477	0.045	0.009	0.462	0.491	27	0.932	0.076	0.482	0.038	0.008	0.469	0.495	24	0.980	0.894
TM2	0.543	0.046	0.009	0.528	0.558	26	0.936	0.109	0.540	0.042	0.008	0.526	0.554	26	0.933	0.089
TM3	0.455	0.020	0.005	0.446	0.463	18	0.982	0.965	0.448	0.021	0.004	0.441	0.456	22	0.905	0.038
TM4	0.446	0.024	0.005	0.437	0.454	23	0.981	0.922	0.450	0.029	0.006	0.440	0.461	23	0.976	0.821
TM5	0.464	0.026	0.006	0.455	0.474	21	0.949	0.331	0.462	0.031	0.006	0.451	0.473	24	0.972	0.709
TM6	0.507	0.047	0.009	0.493	0.522	30	0.970	0.527	0.518	0.048	0.009	0.503	0.533	29	0.967	0.481
TM7	0.445	0.035	0.009	0.430	0.461	16	0.984	0.986	0.441	0.031	0.007	0.428	0.454	17	0.956	0.559
Mean of 7 TMs' mean	0.477	0.036	0.014	0.450	0.503	7	0.865	0.169	0.477	0.038	0.014	0.449	0.505	7	0.871	0.191
Mean of 7 TMs' S.D.	0.035	0.011	0.004	0.026	0.035	7	0.865	0.169	0.034	0.009	0.003	0.028	0.041	7	0.971	0.903
PreMod																
"2"																
TM1	0.421	0.029	0.006	0.410	0.431	22	0.949	0.296	0.426	0.029	0.006	0.415	0.437	21	0.924	0.107
TM2	0.465	0.027	0.005	0.474	0.556	27	0.961	0.382	0.465	0.027	0.005	0.456	0.473	27	0.976	0.771
TM3	0.454	0.021	0.005	0.446	0.463	17	0.983	0.978	0.448	0.021	0.005	0.440	0.456	20	0.903	0.046
TM4	0.447	0.023	0.005	0.439	0.456	22	0.982	0.941	0.450	0.029	0.006	0.440	0.461	23	0.976	0.821
TM5	0.438	0.030	0.007	0.427	0.450	21	0.950	0.336	0.441	0.026	0.006	0.432	0.451	21	0.963	0.587
TM6	0.464	0.036	0.007	0.452	0.475	29	0.970	0.568	0.474	0.037	0.007	0.462	0.481	27	0.971	0.632
TM7	0.481	0.046	0.009	0.465	0.496	27	0.976	0.765	0.478	0.045	0.009	0.463	0.494	26	0.964	0.466
"2 + 1"																
TM1	0.421	0.029	0.006	0.411	0.430	27	0.944	0.150	0.427	0.029	0.006	0.417	0.437	25	0.937	0.125
TM2	0.461	0.032	0.006	0.451	0.472	28	0.942	0.126	0.463	0.027	0.005	0.455	0.472	28	0.982	0.905
TM3	0.455	0.020	0.005	0.446	0.463	18	0.982	0.965	0.448	0.021	0.004	0.441	0.456	22	0.905	0.038
TM4	0.446	0.024	0.005	0.437	0.454	23	0.981	0.922	0.450	0.029	0.006	0.440	0.461	23	0.976	0.821
TM5	0.437	0.030	0.006	0.426	0.448	22	0.947	0.277	0.439	0.026	0.005	0.430	0.449	23	0.955	0.375
TM6	0.464	0.036	0.007	0.452	0.475	29	0.970	0.568	0.470	0.038	0.007	0.458	0.482	29	0.967	0.478
TM7	0.481	0.048	0.008	0.467	0.495	33	0.974	0.589	0.476	0.045	0.008	0.462	0.490	31	0.974	0.622

GPCRs	**Test set 23**								**Test set 24**							
	Accuracy			**90% Confidence Interval for Mean**			**Shapiro-Wilk**		**Accuracy**			**90% Confidence Interval for Mean**			**Shapiro-Wilk**	
	Mean	**S.D.**	**D.E.**	**Lower**	**Upper**	**df**	**StatisticW value**	**Sig. P-value**	**Mean**	**S.D.**	**D.E.**	**Lower**	**Upper**	**df**	**StatisticW value**	**Sig. P-value**
Group 18																
"2"	0.620	0.144	0.054	0.515	0.726	7	0.868	0.179	0.575	0.138	0.052	0.474	0.677	7	0.897	0.313
"2 + 1"	0.681	0.152	0.057	0.569	0.792	7	0.903	0.352	0.647	0.144	0.054	0.541	0.752	7	0.955	0.775
Group 14																
"2"	0.549	0.128	0.048	0.455	0.643	7	0.974	0.927	0.552	0.108	0.041	0.473	0.631	7	0.944	0.676
"2 + 1"	0.605	0.132	0.050	0.508	0.702	7	0.972	0.915	0.620	0.097	0.037	0.549	0.692	7	0.954	0.767

Table 10. The score and the validity of the coding sequences of each trans-membrane segment of 22 GPCRs located in sense chain of chromosome 19 by the scoring matrices of PreMod.

GPCRs		Score							Validity							Accumulation			
Samples	contig	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	"2"	"1"	"2 + 1"	"0"
threshold		0.417	0.455	0.396	0.401	0.430	0.458	0.462											
OR4F17_HUMAN	1								0	0	0	0	0	0	0	0	0	0	7
KISSR_HUMAN	1	0.463					0.495	0.502	2	0	0	0	0	2	2	3	0	3	4
EDG6_HUMAN (S1PR4_HUMAN)	1	0.418		0.444			0.494	0.535	2	0	2	0	0	2	2	4	0	4	3
EMR1_HUMAN	1			0.397					0	0	2	0	0	0	0	1	0	1	6
OR2Z1_HUMAN	3	0.450				0.433			2	0	0	0	2	0	0	2	0	2	5
OR1M1_HUMAN	3	0.458			0.406				2	0	0	2	0	0	0	2	0	2	5
OR7D2_HUMAN	3	0.459						0.469	2	0	0	0	0	0	2	2	0	2	5
P2Y11_HUMAN	3	0.421		0.399	0.423				2	0	2	2	0	0	0	3	0	3	4
OR7C2_HUMAN	3	0.449	0.459					0.485	2	2	0	0	0	0	0	2	0	2	5
OR1I1_HUMAN	3	0.436						0.467	2	0	0	0	0	0	2	2	0	2	5
O10H2_HUMAN	3	0.461		0.398					2	0	2	0	0	0	0	2	0	2	5
O10H3_HUMAN	3	0.449				0.431 0.443			2	0	0	0	1 1	0	0	1	1	2	5
O10H5_HUMAN	3	0.461		0.398					2	0	2	0	0	0	0	2	0	2	5
O10H4_HUMAN	3	0.459							2	0	0	0	0	0	0	1	0	1	6
PAR4_HUMAN	3	0.439	0.481			0.437		0.529	2	2	0	0	2	0	2	4	0	4	3
FFAR1_HUMAN	4	0.441	0.470		0.413	0.439			2	2	0	2	1	0	0	3	1	4	3
FFAR3_HUMAN	4	0.423	0.473	0.432		0.431	0.490	0.501	2	2	2	0	1	2	2	5	1	6	1
FFAR2_HUMAN	4	0.436	0.483			0.454		0.463	2	2	0	0	1	0	2	3	1	4	3
GIPR_HUMAN	4								0	0	0	0	0	0	0	0	0	0	7
C5AR_HUMAN	4	0.497	0.478	0.468 0.395	0.412		0.494	0.572	2	2	2 1	2	0	2	2	6	0	6	1
C5ARL_HUMAN	4	0.475	0.476	0.450	0.412		0.516	0.486	2	2	1	2	0	2	2	5	1	6	1
GPR32_HUMAN	4	0.432	0.476	0.402			0.483	0.540	2	2	2	0	0	2	2	5	0	5	2
"2"									19/22	8/22	8/22	5/22	2/22	6/22	10/22	20/22			
"2 + 1"									19/22	8/22	9/22	5/22	6/22	6/22	10/22			20/22	
"0"																			2/22

of molecular similarities required on the order of a million individual protein segment/segment similarities. Rather than employ the phylogenetic evolution similarity method directly, we employed the scoring matrix approach to infer similarities in these protein sequences. Our previous work focused on the use of this technique for computations involving building scoring matrices for prediction. Consequently, the following work was wanted to verify that the method yielded the expected results within the space of known GPCRs in test sets.

TM1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	0.003	0.001	0.001	0.002	0.006	0.001	0.003	0.001	0.002	0.007	0	0.001	0.006	0.001	0.004	0.005	0.001	0.003	0.003	0
T	0.003	0.016	0.004	0.013	0.009	0.005	0.009	0.009	0.004	0.002	0.013	0.004	0.002	0.016	0.001	0.012	0.011	0.009	0.002	0.017
C	0.007	0.003	0.008	0.002	0.002	0.01	0.002	0.004	0.007	0.004	0.006	0.011	0.005	0.004	0.01	0.001	0.004	0.007	0.009	0.003
G	0.008	0.001	0.008	0.004	0.003	0.005	0.008	0.006	0.008	0.008	0.003	0.006	0.008	0	0.006	0.003	0.004	0.002	0.007	0.001
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
A	0.001	0.005	0	0.002	0.006	0	0.003	0.004	0	0	0.003	0	0.002	0.004	0.004	0.003	0.017	0.016	0.002	0.006
T	0.003	0.002	0.017	0.004	0.002	0.004	0.004	0.003	0.018	0.004	0.007	0.017	0.004	0	0.003	0.002	0.001	0.001	0.009	0.003
C	0.004	0.006	0.004	0.006	0.001	0.006	0.01	0.006	0.003	0.009	0.007	0.003	0.005	0.002	0.004	0.009	0	0.003	0.009	0.004
G	0.013	0.008	0	0.009	0.012	0.011	0.004	0.007	0	0.008	0.004	0.001	0.009	0.015	0.01	0.006	0.003	0.001	0.001	0.008
	41	42	43	44	45	46	47	48												
A	0	0.002	0.004	0.001	0.005	0.007	0.001	0												
T	0.013	0.001	0.004	0.017	0.002	0.002	0.016	0.007												
C	0.004	0.013	0.009	0.003	0.007	0.001	0.003	0.008												
G	0.004	0.004	0.004	0.001	0.007	0.011	0.001	0.006												

TM2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	0.011	0.001	0.003	0.006	0.001	0.002	0.015	0.014	0.001	0.001	0	0	0.004	0	0.001	0.003	0.001	0.002	0.004	0.001
T	0.006	0.021	0.007	0.005	0.015	0.004	0.007	0.001	0.004	0.005	0.026	0.004	0.007	0.004	0.007	0.008	0.019	0.007	0.006	0.003
C	0.005	0.003	0.01	0.007	0.004	0.012	0.002	0.005	0.018	0.019	0	0.006	0	0.018	0.014	0.008	0.001	0.006	0	0.018
G	0.004	0	0.007	0.008	0.0004	0.007	0.002	0.005	0.001	0	0	0.015	0.015	0.004	0.004	0.007	0.004	0.01	0.015	0.004
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	
A	0.008	0.003	0.023	0.001	0.004	0	0.001	0.002	0	0.004	0.004	0.004	0.004	0.007	0	0.002	0.009	0	0.002	
T	0.007	0.001	0	0.007	0.005	0.023	0.004	0.005	0.02	0.005	0.008	0.017	0.005	0.005	0.009	0.005	0.006	0.017	0.007	
C	0.009	0.001	0.002	0.017	0.012	0.001	0.007	0.012	0.004	0.008	0.004	0.004	0.009	0.004	0.01	0.012	0.004	0.004	0.012	
G	0.002	0.021	0.001	0.001	0.004	0.001	0.014	0.006	0.002	0.008	0.01	0.001	0.008	0.009	0.007	0.006	0.007	0.004	0.004	

TM3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	0.01	0.006	0.001	0.009	0	0.001	0.007	0.009	0.004	0.005	0.001	0.003	0.009	0.001	0.003	0.014	0	0.002	0.001	0.002
T	0.011	0.005	0.003	0.007	0.016	0.001	0.011	0.009	0.006	0.002	0.005	0.003	0.013	0.001	0.004	0	0.022	0.003	0.013	0.02
C	0.001	0.009	0.02	0.003	0.003	0.01	0.004	0.006	0.009	0.001	0.014	0.016	0.001	0.016	0.016	0.001	0.003	0.014	0.009	0.001
G	0.003	0.006	0.002	0.007	0.006	0.014	0.003	0.002	0.006	0.018	0.006	0.003	0.002	0.008	0.002	0.01	0.001	0.006	0.003	0.002
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	
A	0.001	0.013	0.006	0.001	0.004	0.001	0.002	0.008	0	0.003	0.002	0	0.001	0.017	0	0.001	0.013	0.001	0.002	
T	0.004	0.006	0.01	0.008	0.002	0.024	0.002	0.006	0.008	0.008	0.004	0.01	0.001	0	0.024	0.007	0.003	0	0.008	
C	0.012	0.006	0.005	0.011	0.016	0	0.009	0.004	0.009	0.008	0.003	0.01	0.019	0.006	0.001	0.012	0.001	0.013	0.015	
G	0.009	0.001	0.004	0.006	0.003	0.001	0.012	0.007	0.008	0.007	0.017	0.005	0.004	0.002	0	0.006	0.009	0.012	0.001	

TM4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	0.005	0	0.003	0.006	0.001	0.003	0.014	0	0.005	0	0	0.001	0.005	0	0.002	0.007	0.001	0.004	0.004	0
T	0.004	0.011	0.004	0.003	0.014	0.006	0.002	0.023	0.001	0.027	0.001	0.001	0.002	0.018	0.004	0.004	0.026	0.005	0.014	0.004

C	0.005	0.012	0.014	0.006	0.006	0.017	0.002	0.005	0.017	0.001	0.002	0.001	0.004	0.008	0.014	0.012	0.003	0.009	0.003	0.024	
G	0.017	0.008	0.01	0.014	0.009	0.004	0.012	0.002	0.006	0.001	0.027	0.027	0.019	0.004	0.009	0.007	0.001	0.013	0.01	0.002	
	21	22	23	24	25	26	27	28	29	30	31	32	33								
A	0.003	0.008	0.001	0.004	0.004	0	0.003	0.009	0.001	0	0.01	0.001	0.005								
T	0.005	0.006	0.022	0.004	0.004	0.017	0.001	0.008	0.021	0.005	0.009	0.005	0.003								
C	0.014	0.009	0.006	0.009	0.008	0.009	0.017	0.004	0.006	0.017	0.001	0.019	0.018								
G	0.009	0.007	0.001	0.013	0.014	0.004	0.009	0.009	0.002	0.009	0.011	0.005	0.004								

TM5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
A	0.008	0	0.002	0.006	0.002	0.004	0.001	0.003	0.002	0.008	0.004	0.003	0.007	0.001	0.003	0.006	0.005	0.005	0.003	0	
T	0.006	0.02	0.002	0.008	0.008	0.004	0.02	0.023	0.004	0.008	0.018	0.005	0.004	0.02	0.005	0.001	0.005	0.008	0.009	0.026	
C	0.006	0.005	0.014	0.004	0.009	0.017	0.005	0.001	0.018	0.007	0.005	0.012	0.007	0.004	0.007	0.017	0.016	0.011	0.014	0.002	
G	0.008	0.002	0.009	0.01	0.008	0.002	0.002	0.001	0.003	0.005	0.001	0.007	0.01	0.002	0.012	0.004	0.002	0.003	0.002	0	
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36					
A	0.002	0.004	0.001	0.002	0.011	0.001	0.001	0.016	0	0.001	0.007	0.001	0.002	0.006	0.003	0.001					
T	0.007	0.008	0.018	0.007	0.004	0.024	0.004	0.007	0.027	0.006	0.007	0.016	0.016	0.01	0.019	0.008					
C	0.007	0.005	0.007	0.011	0.007	0.003	0.018	0.003	0.001	0.012	0.005	0.011	0.011	0.004	0.004	0.011					
G	0.011	0.011	0.002	0.007	0.006	0	0.005	0.002	0	0.009	0.008	0.001	0.001	0.008	0.002	0.007					

TM6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
A	0.01	0	0.002	0.003	0	0.003	0.006	0	0.003	0.001	0.002	0	0.01	0	0.005	0.009	0	0.002	0.004	0	
T	0.001	0.022	0.008	0.006	0.014	0.005	0.002	0.017	0.006	0.021	0.026	0.011	0.002	0.017	0.005	0.002	0.021	0.003	0.021	0.007	
C	0.008	0.005	0.006	0.006	0.005	0.011	0.005	0.01	0.011	0.003	0	0.013	0.004	0.008	0.013	0.007	0.006	0.02	0.002	0.005	
G	0.01	0.001	0.013	0.013	0.009	0.009	0.015	0.002	0.008	0.002	0	0.003	0.012	0.002	0.006	0.01	0.001	0.002	0.001	0.016	
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36					
A	0.002	0	0.002	0.001	0.002	0	0.004	0.002	0	0.004	0.005	0.012	0.001	0.008	0.013	0					
T	0.002	0.025	0.007	0.004	0.006	0.014	0.002	0	0.006	0.006	0.013	0.013	0.009	0.009	0.011	0.008					
C	0.021	0.001	0.001	0.005	0.01	0.006	0.01	0.024	0.022	0.015	0.006	0	0.016	0.01	0.002	0.014					
G	0.003	0.002	0.017	0.018	0.01	0.007	0.012	0.002	0	0.002	0.003	0.003	0.002	0.002	0.002	0.006					

TM7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
A	0.005	0.002	0.002	0.007	0.002	0.004	0.009	0.011	0	0.006	0.001	0.004	0.019	0.019	0.002	0.008	0.001	0.001	0.006	0	
T	0.006	0.021	0.007	0.003	0.005	0.004	0.011	0.006	0.007	0.009	0.018	0.004	0.001	0.001	0.008	0.01	0.001	0.006	0.008	0.012	
C	0.014	0.003	0.01	0.002	0.01	0.014	0.002	0.004	0.015	0.004	0.004	0.016	0.006	0.004	0.012	0.003	0.008	0.016	0.016	0.006	
G	0	0	0.007	0.014	0.009	0.003	0.004	0.004	0.004	0.006	0.003	0.003	0	0.002	0.004	0.005	0.016	0.003	0.007	0.007	
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39		
A	0.001	0.008	0	0.002	0.017	0.019	0.001	0	0	0.005	0.011	0	0.002	0.011	0.001	0.002	0.001	0.022	0		
T	0.004	0.006	0.024	0.002	0	0.005	0.007	0.002	0.002	0.007	0.004	0.024	0.004	0.004	0.022	0.007	0.024	0.002	0.01		
C	0.012	0.007	0.002	0.013	0	0	0.015	0.02	0.023	0.011	0.006	0.002	0.006	0.006	0.001	0.001	0.001	0.002	0.015		
G	0.008	0.004	0	0.009	0.009	0.001	0.003	0.004	0.001	0.003	0.005	0	0.007	0.004	0.002	0.005	0	0	0.001		

Figure 8. The score matrix of the prediction model.

4. DISCUSSION

Being the largest family of cell surface receptors, GPCRs play a key role in cellular signaling pathways that regulate many basic physiological processes, such as neurotransmission, secretion, growth, cellular differentiation, inflammatory, and immune responses [46,47]. Protein phosphorylation is an essential type of post-translational modification that consists of the addition of a phosphate group to serine (S), threonine (T), and tyrosine (Y) [48]. This process is catalyzed by a group of enzymes called kinases, and can be reversed by phosphatases [47]. The phosphorylation process is catalyzed by GPCR kinases (GRKs) that recognize the receptors as substrates after agonist binding [49]. This phosphorylation often modifies the cytosolic C-terminal tail and leads to receptor uncoupling from G proteins, binding arrestin, and further results in receptor desensitization and deactivation [50]. Huang JH and co-work have revealed that the exact positions of a phosphorylation in a GPCR protein sequence could provide useful clues for drug design and other biotechnology applications [47].

GPCRs are extensively targeted for drug development in humans, especially the biogenic amine-binding GPCRs, which are integral components of the central and peripheral nervous systems of eukaryotes and include receptors that bind the neurotransmitters dopamine, histamine, octopamine, serotonin, tyramine, and acetylcholine [51, 52]. Malaria is a devastating infection caused by protozoa of the genus Plasmodium (*P. falciparum*). Gamo *et al.* have reported multiple GPCR-interacting chemistries as promising anti-malarial leads [53]. Analyses using historic assay data revealed that some compounds had activity, but against drug targets without obvious orthologues in the malarial genome, such as GPCRs, nuclear recaptors, ion channels and transporters. They suggested several novel mechanisms of antimalarial action, such as inhibition of protein kinases and host-pathogen interactiton related targets, which provide new tools to exploit the malarial kinome for drug discovery [53]. More than 100 different GPCRs have been identified in the genomes of multiple insect species, including malaria- and yellow fever-transmitting mosquitoes. Hill *et al.* used bioinformatics approaches to identify a total of 276 GPCRs from the Anopheles gambiae genome, which are likely to play roles in pathways affecting almost every aspect of the mosquito's life cycle [54]. Meyer JM *et al.* used "genome-to-lead" approach to develop new mode-of-action insecticides for arthropod disease vectors, involving 1) exploitation of an arthropod genome sequence for novel target identification; 2) molecular, biochemical and pharmacological target validation; 3) chemical library screening; and 4) confirmation of hits and identification of candidate "leads" using secondary *in vitro* assays and mosquito *in vivo* assays [52]. They reported the first study to identify Aedes aegypti D1-like dopamine receptor (AaDOP1) antagonists with *in vivo* toxicity toward mosquitoes.

GPCRs comprise the largest family of validated drug targets while 30% - 50% of approved drugs derive their benefits by selective targeting of GPCRs [55]. Mutations in GPCRs are responsible for over 30 disorders, includeing cancers, heritable obesity, diabetes insipidus, blindness, endocrine diseases, and diseases involving the melanocortin type 4 and gonadotropin releasing hormone receptor (GnRHR) [56,57]. Many pathologies associated with misfolded mutant receptors occur because these are retained by the endoplasmic reticulum (ER) and do not reach their normal site of function [57]. Normally, GPCRs are subjected to a stringent quality control system (QCS) in the endoplasmic reticulum [56]. This system consists of both protein chaperones and enzyme-like proteins. The former retains misfolded proteins while the latter participates in catalysis of the folding process. Moreover, the QCS insures that only correctly folded proteins enter the pathway leading to the plasma membrane. However, point mutations may result in the production of misfolded and disease-causing proteins that are unable to reach their functional destinations in the cell because they are retained by the QCS even though they may retain function [56]. On the other hand, pharmacoperone drugs (from "pharmacological chaperone") are small molecules that enter cells and serve as a "molecular scaffold" to promote correct folding of otherwise-misfolded mutant proteins and route correctly within the cell [58]. Because these drugs are frequently selected from candidates that were originally identified as target specific antagonists, they also show high target specificity as pharmacoperones, although competition for endogenous ligands is a therapeutic complication. Accordingly Janovick *et al.* sought to develop assays that would identify molecules that were not necessarily agonists or antagonists [56]. In principle, the pharmacoperone-rescue approach applies to a diverse array of human diseases that result from protein misfolding, such as cystic fibrosis [59], hypogonadotropic hypogonadism [60], nephrogenic diabetes insipidus [61], retinitis pigmentosa [62], hyper-cholesterolemia [63], cataracts [64], neurodegenerative diseases (Huntington's [65], Alzheimer's [66], Parkinson's [67]) and particular cancers [68]. Janovick *et al.* have also explored molecular mechanism of action of pharmacoperone rescue of misrouted GPCR mutants using hGnRHR, a useful model for studying pharmacoperones [57]. Especially, there is a naturally occurring and highly conserved salt bridge (E^{90}-K121) in hGnRHR that stabilizes the relation between transmembranes 2 and 3 of hGnRHR, which is required for passage of the receptor through the cellular QCS and to

the plasma membrane. This bridge, broken in the naturally occurring hGnRHR mutant $E^{90}K$, causes hypogonadotropic hypogonadism because the misfolded mutant receptor fails the cellular QCS and cannot traffic to the plasma membrane [69]. Additionally, pharmacoperone drugs from different chemical classes all happened to interact identically by creating a surrogate bridge for E^{90}-K121. This ligand-mediated bridge plays a key role in rescue of misrouted GPCR mutants. The method provides the basis of novel primary screens for pharmacoperones, especially to identify structures beyond agonists or antagonists. Non-antagonistic pharmacoperones have a therapeutic advantage since they will not compete for endogenous agonists and may not have to be washed out once rescue has occurred and before activation by endogenous or exogenous agonists [56]. These studies suggest that rational design of these therapeutic agents, e.g. ones that do not compete with endogenous ligands, is likely to assist this therapeutic approach.

GPCRs are a large superfamily of membrane bound signaling proteins that are involved in the regulation of a wide range of physiological functions and constitute the most common target for therapeutic intervention [70]. GPCRs are among the most important drug targets for the pharmaceutical industry. Knowledge of the three-dimensional structure of a protein is of utmost importance for drug discovery, as it serves as the basis for the identification of novel ligands by means of computational or *in silico* techniques, such as de novo design and virtual screening. 25% of the small molecule drugs approved in 2006 were discovered through structure-based drug discovery (SBDD) [70]. Consequently, target identification is a critical step following the discovery of small molecules that elicit a biological phenotype. There are a serial of technologies and approaches applied in new drug targets and biomarker identification, such as proteomics technology, systems biology approach, microRNA technology, and computational methods. Sugahara *et al.* have identified a large number of candidates for the target proteins specific to β1,4-galactosyltransferase-I (β4GalT-I) by comparative analysis of β4-GalT-I-deleted and wild-type mice using the LC/MS-based technique with the isotope-coded glycosylation site-specific tagging (IGOT) of lectin-captured N-glycopeptides [71]. Their approach to identify the target proteins in a proteome-scale offers common features and trends in the target proteins, which facilitate understanding of the mechanism that controls assembly of a particular glycan motif on specific proteins. Research on microRNAs (miRNAs) is a promising new research, providing novel insights into the pathogenesis of some diseases, biomarker identification, and treatment. The short (approximately 22 nucleotides), endogenous, widely distributed, single-stranded RNAs target both Mrna degradation and suppression of protein translation based on sequence complementarity between the miRNA and its targeted mRNA [72]. During evolution, RNA retroviruses or transgenes invaded the eukaryotic genome and inserted itself in the noncoding regions of DNA, acting as transposon-like jumping genes. MiRNAs are evolutionary conserved in animals and plants, and regulate specific target mRNAs at the post-transcriptional level, which involved in several biological processes, including development, cell differentiation, proliferation and apoptosis [73]. MiRNAs may be responsible for regulating the expression of nearly one-third of the genes in the human genome whereas very little is known about their biological functions and functional targets despite the identification of more than 1900 mature human miRNAs. Furthermore, miRNA deregulation often results in an impaired cellular function, and a disturbance of downstream gene regulation and signaling cascades, suggesting their implication in disease etiology. Koskun M *et al.* have identified dysregulated miRNAs in tissue samples of inflammatory bowel disease (IBD) patients, demonstrated similar differences in circulating miRNAs in the serum of IBD patients, and further discovered that miRNAs will aid in the early diagnosis of IBD and in the development of personalized therapies [73]. Additionally, our results represent a generalization of the validation and identification of GPCRs using computational methods. The sequence similarity and protein diversity exhibited intuitive behavior in the clustering when considering the underlying distributions. The computations involving the scoring matrix methods are a substantial test of such an approach, with explicit models built that cover roughly 90% of approved GPCRs in test sets. Our focus in previous work was methodological for prediction of active sites using the scoring matrix [40] while this approach showed that the scoring matrix methodology quantitatively outperformed molecular modeling methods for prediction target proteins. In the present work, the scoring matrix methods are used to predict potential proteins as well as prediction of active sites at a level of genome or amino acid sequences.

In conclusion, the present work seeks to provide an in silico model of known GPCR protein fishing technologies in order to rapidly fish out potential drug targets on the basis of amino acid sequences and seven TMs of GPCRs. Some scoring matrices were trained on 22 groups of GPCRs in the GPCRDB database. These models were employed to predict the GPCR proteins in two groups of test sets. On average, the mean correct rate of each TM of 38 GPCRs from T23 and T24 was found 62% and 57.5%, respectively, using training set 18 $\left(S_{D18}^{L}\right)$; the mean hit rate of each TM of 38 GPCRs from T23 and T24 was found 68.1% and 64.7%, respectively. Based on the scoring matrices of PreMod, the

mean cor- rect rate of each TM of GPCRs from T23 and T24 was found 62% and 62.04%, respectively; the mean hit rate of each TM of GPCRs from T23 and T24 was found 67.7% and 68.0%, respectively. The means of GPCRs in Test set 23 based on S_{D18}^{L} is close to those based on PreMod; whereas the means of GPCRs in Test set 24 based on S_{D18}^{L} is less than those based on PreMod. Moreover, the S_{D18}^{L} accuracy ("2") and validity ("2 + 1") rates of prediction all seven TMs of 38 GPCRs by the scoring matrices of PreMod are more than those by S_{D18}^{L}, S_{A14}^{L} and S_{A3}^{L}; whereas the hit rates (94.74% and 97.37%) by PreMod are less than those of S_{A3}^{L} but bigger than those of S_{D18}^{L} and S_{A14}^{L}. This is the reason that we choose PreMod to predict some potential drug targets. 23 GPCR proteins in the sense chain of chromosome 19 constructing validation set were predicted and validated by PreMod whose hit rate is up to 95.65%. Further evaluation is under investigation.

5. ACKNOWLEDGEMENTS

This work was supported by a grant from Basic Scientific Research Expenses of Central University (020814360012) and National Key Technology R&D Program (2008BAI51B01).

REFERENCES

[1] Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nature Reviews Drug Discovery*, **1**, 727-730. doi:/10.1038/nrd892

[2] Neves, S.R., Ram, P.T. and Iyengar, R. (2002) G protein pathways. *Science*, **296**, 1636-1639. doi:/10.1126/science.1071550

[3] Hynes, R.O. (2002) Integrins: Bidirectional, allosteric signaling machines. *Cell*, **110**, 673-687. doi:/10.1016/S0092-8674(02)00971-6

[4] Tilley, D.G. (2011) G protein-dependent and G protein-independent signaling pathways and their impact on cardiac function. *Circulation Research*, **109**, 217-230. doi:/10.1161/CIRCRESAHA.110.231225

[5] Shen, B., Delaney, M.K. and Du, X. (2012) Inside-out, outside-in, and inside-outside-in: G protein signaling in integrin-mediated cell adhesion, spreading, and retraction. *Current Opinion in Cell Biology*, **24**, 600-606. doi:/10.1016/j.ceb.2012.08.011

[6] Jaffe, A.B. and Hall, A. (2005) Rho GTPases: Biochemistry and biology. *Annual Review of Cell and Developmental Biology*, **21**, 247-269. doi:/10.1146/annurev.cellbio.21.020604.150721

[7] Bockaert, J. and Pin, J.P. (1999) Molecular tinkering of G protein-coupled receptors:an evolutionary success. *The EMBO Journal*, **18**, 1723-1729. doi:/10.1093/emboj/18.7.1723

[8] Vassilatis. D.K., Hohmann, J.G., Zeng, H., Li, F., Ranchalis, J.E., Mortrud, M.T., Brown, A., Rodriguez, S.S., Weller, J.R., Wright, A.C., Bergmann, J.E. and Gaitanaris, G.A. (2003) The G protein-coupled receptor repertoires of human and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 4903-4908. doi:/10.1073/pnas.0230374100

[9] Foord, S.M., Bonner, T.I., Neubig, R.R., Rosser, E.M., Pin, J.P., Davenport, A.P., Spedding, M. and Harmar, A.J. (2005) International Union of Pharmacology. XLVI. G protein-coupled receptor list. *Pharmacological Reviews*, **57**, 279-288. doi:/10.1124/pr.57.2.5

[10] Horn, F., Weare, J., Beukers, M.W., Hörsch, S., Bairoch, A., Chen, W., Edvardsen, O., Campagne, F. and Vriend, G. (1998) GPCRDB: An information system for G protein-coupled receptors. *Nucleic Acids Research*, **26**, 275-279. doi:/10.1093/nar/26.1.275

[11] Attwood, T.K. and Findlay, J.B. (1994) Fingerprinting G-protein-coupled receptors. *Protein Engineering*, **7**, 195-203. doi:/10.1093/protein/7.2.195

[12] Harmar, A.J. (2001) Family-B G-protein-coupled receptors. *Genome Biology*, **2**, 3013.1-3013.10. doi:/10.1186/gb-2001-2-12-reviews3013

[13] Bräuner-Osborne, H., Wellendorph, P. and Jensen, A.A. (2007) Structure, pharmacology and therapeutic prospects of family C G-protein coupled receptors. *Current Drug Targets*, **8**, 169-184. doi:/10.2174/138945007779315614

[14] Herskowitz, I. and Marsh, L. (1988) STE2 protein of Saccharomyces kluyveri is a member of the rhodopsin/beta-adrenergic receptor family and is responsible for recognition of the peptide ligand alpha factor. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 3855-3859. doi:/10.1073/pnas.85.11.3855

[15] Devreotes, P.N., Kimmel, A.R., Johnson, R.L., Klein, P.S., Sun, T.J. and Saxe III, C.L. (1988) A chemoattractant receptor controls development in Dictyostelium discoideum. *Science*, **241**, 1467-1472. doi:/10.1126/science.3047871

[16] Malbon, C.C. (2004) Frizzleds: New members of the superfamily of G-protein-coupled receptors. *Frontiers in Bioscience*, **9**, 1048-1058. doi:/10.2741/1308

[17] Taipale, J., Chen, J.K., Cooper, M.K., Wang, B., Mann, R.K., Milenkovic, L., Scott, M.P. and Beachy, P.A. (2000) Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature*, **406**, 1005-1009. doi:/10.1038/35023008

[18] Buck, L. and Axel, R. (1991) A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell*, **65**, 175-187. doi:/10.1016/0092-8674(91)90418-X

[19] Mombaerts, P. (1999) Seven-transmembrane proteins as odorant and chemosensory receptors. *Science*, **286**, 707-711. doi:/10.1126/science.286.5440.707

[20] Firestein, S. (2000) The good taste of genomics. *Nature* **404**, 552-553. doi:/10.1038/35007167

[21] Howard, A.D., McAllister, G., Feighner, S.D., Liu, Q., Nargund, R.P., Van der Ploeg, L.H. and Patchett, A.A. (2001) Orphan *G*-protein-coupled receptors and natural ligand discovery. *Trends in Pharmacological Sciences*, **22**, 132-140.

doi:/10.1016/S0165-6147(00)01636-9

[22] Lee, D.K., George, S.R., Evans, J.F., Lynch, K.R. and O' Dowd, B.F. (2001) Orphan G protein-coupled receptors in the CNS. *Current Opinion in Pharmacology*, **1**, 31-39. doi:/10.1016/S1471-4892(01)00003-0

[23] Filmore, D. (2004) It's a GPCR world. *Modern Drug Discovery*, **11**, 24-28.

[24] Wise, A., Gearing, K. and Rees, S. (2002) Target validation of G-protein coupled receptors. *Drug Discovery Today*, **7**, 235-246. doi:/10.1016/S1359-6446(01)02131-6

[25] Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001) The sequence of the human genome. *Science*, 1304-1351. doi:/10.1126/science.1058040

[26] Meneses, A. (1999) 5HT system and cognition. *Neuroscience & Biobehavioral Reviews*, **23**, 1111-1125. doi:/10.1016/S0149-7634(99)00067-6

[27] Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P. and McKernan, K. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921. doi:/10.1038/35057062

[28] O'Dowd, B.F., Ji, X., Alijaniaram, M., Nguyen, T. and George S.R. (2006) A novel drug screening assay for G protein-coupled receptors. In: Rognan, D., Mannhold, R., Kubinyi, H. and Folkers, G., Eds., *Ligand Design for G Protein-Coupled Receptors* (*Methods and Principles in Medicinal Chemistry*), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 51-60.

[29] Baldwin, J.M. (1993) The probable arrangement of the helices in G protein-coupledreceptors. *The EMBO Journal*, **12**, 1693-1703.

[30] Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M. and Henderson, R. (1996) Electron-crystallographic refinement of the structure of bacteriorhodopsin. *Journal of Molecular Biology*, **259**, 393-421. doi:/10.1006/jmbi.1996.0328

[31] Kimura, Y., Vassylyev, D.G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T. and Fujiyoshi, Y. (1997) Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature*, **389**, 206-211. doi:/10.1038/38323

[32] Pebay-Peyroula, E., Rummel, G., Rosenbusch, J.P. and Landau, E.M. (1997) X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science*, **277**, 1676-1681. doi:/10.1126/science.277.5332.1676

[33] Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Trong, I.L., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M. and Miyano, M. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, **289**, 739-745. doi:/10.1126/science.289.5480.739

[34] Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R., Sanishvili, R., Fischetti, R.F., Schertler, G.F., Weis, W.I. and Kobilka, B.K. (2007) Crystal structure of the human β_2-adrenergic G-protein-coupled receptor. *Nature*, **450**, 383-387. doi:/10.1038/nature06325

[35] Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K. and Stevens, R.C. (2007) High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science*, **318**, 1258-1265. doi:/10.1126/science.1150577

[36] Drews, J. (1996) Genomic sciences and the medicine of tomorrow. *Nature Biotechnology*, **14**, 1516-1518. doi:/10.1038/nbt1196-1516

[37] Yang, J. and Liu, C.Q. (2000) Molecular modeling on human CCR5 receptors and complex with CD4 antigens and HIV-1 envelope Glycoprotein gp120. *Acta Pharmacologica Sinica*, **20**, 29-34.

[38] Yang, J., Zhang, Y.W., Huang, J.F., Zhang, Y.P. and Liu, C.Q. (2000) Structure analysis of CCR5 from human and primates. *Journal of Molecular Structure*: *Theochem*, **505**, 199-210. doi:/10.1016/S0166-1280(99)00393-0

[39] Yang, J. and Hua, W.Y. (1996). Basic pharmacophore for some antithrombotic agents with combined thromboxane receptor antagonists (TXRA)/thromboxine synthase inhibitor (TXSI) activities. *Drug Development Research*, **39**, 197-200. doi:/10.1002/(SICI)1098-2299(199610)39:2<197::AID-DDR14>3.0.CO;2-9

[40] Zhan, C.Y., Yang, J., Dong, X.C. and Wang, Y.L. (2007) Molecular modeling of purinergic receptor P2Y12 and interaction with its antagonists. *Journal of Molecular Graphics and Modelling*, **26**, 20-31. doi:/10.1016/j.jmgm.2006.09.006

[41] Xiao, Y.D., Harris, R., Bayram, E., Santago II, P. and Schmitt, J.D. (2006) Supervised self-organizing maps in drug discovery. 2. Improvements in descriptor selection and model validation. *Journal of Chemical Information and Modeling*, **46**, 137-144. doi:/10.1021/ci0500841

[42] Yang, J., Dong, X.C. and Leng, Y. (2006) Application of FTTP to alpha-helix or beta-strand motifs. *Journal of Theoretical Biology*, **242**, 199-219. doi:/10.1016/j.jtbi.2006.02.014

[43] Yang, J., Dong, X.C. and Leng, Y. (2006) Conformation biases of amino acids based on tripeptide microenvironment from PDB database. *Journal of Theoretical Biology*, **240**, 374-384. doi:/10.1016/j.jtbi.2005.09.025

[44] Yu, J.M., Li, D.D., Xu, Z.Q., Cheng, W.X., Zhang, Q., Li, H.Y., Cui, S.X., Yang, M.-J., S.H., Fang, Z.Y. and Duan, Z.J. (2008) Human bocavirus infection in children hospitalized with acute gastroenteritis in China. *Journal of Clinical Virology*, **42**, 280-285. doi:/10.1016/j.jcv.2008.03.032

[45] Steiger, N.M., Lada, E.K., Wilson, J.R., Alexopoulos, C., Goldsman, D. and Zouaoui, F. (2002) ASAP2: An Improved batch means procedure for simulation output analysis. In: Yücesan, E., Chen, C.-H., Snowdon, J.L. and Charnes, J.M., Eds., *Proceedings of the* 2002 *Winter Simulation Conference*, Piscataway, New Jersey, 336-344.

[46] Attwood, T.K., Croning, M.D. and Gaulton, A. (2002) Deriving structural and functional insights from a ligand-based hierarchical classification of G protein-coupled receptors. *Protein Engineering*, **15**, 7-12. doi:/10.1093/protein/15.1.7

[47] Huang, J.H., Cao, D.S., Yan, J., Xu, Q.S., Hu, Q.N. and Liang, Y.Z. (2012) Using core hydrophobicity to identify phosphorylation sites of human G protein-coupled receptors. *Biochimie*, **94**, 1697-1704. doi:/10.1016/j.biochi.2012.03.022

[48] Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912-1934. doi:/10.1126/science.1075762

[49] Pitcher, J.A., Freedman, N.J. and Lefkowitz, R.J. (1998) G protein-coupled receptor kinases. *Annual Review of Biochemistry*, **67**, 653-692. doi:/10.1146/annurev.biochem.67.1.653

[50] Tobin, A.B., Butcher, A.J. and Kong, K.C. (2008) Location, location, location site-specific GPCR phosphorylation offers a mechanism for cell-type-specific signaling. *Trends in Pharmacological Sciences*, **29**, 413-420. doi:/10.1016/j.tips.2008.05.006

[51] Hauser, F., Cazzamali, G., Williamson, M., Blenau, W. and Grimmelikhuijzen, J.P. (2006) A review of neurohormone GPCRs present in the fruitfly *Drosophila melanogaster* and the honey bee *Apis mellifera*. *Progress in Neurobiology*, **80**, 1-19. doi:/10.1016/j.pneurobio.2006.07.005

[52] Meyer, J.M., Ejendal, K.F., Avramova, L.V., Garland-Kuntz, E.E., Giraldo-Calderón, G.I., Brust, T.F., Watts, V.J. and Hill, C.A. (2012) A "genome-to-lead" approach for insecticide discovery: Pharmacological characterization and screening of *Aedes aegypti* D(1)-like dopamine receptors. *PLOS Neglected Tropical Diseases* **6**, e1478. doi:/10.1371/journal.pntd.0001478

[53] Gamo, F.J., Sanz, L.M., Vidal, J., de Cozar, C., Alvarez, E., Lavandera, J.L., Vanderwall, D.E., Green, D,V., Kumar, V., Hasan, S., Brown, J.R., Peishoff, C.E., Cardon, L.R. and Garcia-Bustos, J.F. (2010) Thousands of chemical starting points for antimalarial lead identification. *Nature*, **465**, 305-312. doi:/10.1038/nature09107

[54] Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., Chrystal, M.A., Cravchik, A., Collins, F.H., Robertson, H.M. and Zwiebel, L.J. (2002) G protein-coupled receptors in *Anopheles gambiae*. *Science*, **298**, 176-178. doi:/10.1126/science.1076196

[55] Gruber, C.W., Muttenthaler, M. and Freissmuth, M. (2010) Ligand-based peptide design and combinatorial peptide libraries to target G protein-coupled receptors. *Current Pharmaceutical Design*, **16**, 3071-3088. doi:/10.2174/138161210793292474

[56] Janovick, J.A., Park, B.S. and Conn, P.M. (2011) Therapeutic rescue of misfolded mutants: Validation of primary high throughput screens for identification of pharmacoperone drugs. *PLoS One*, **6**, e22784. doi:/10.1371/journal.pone.0022784

[57] Janovick, J.A., Patny, A., Mosley, R., Goulet, M.T., Altman, M.D., Rush 3rd, T.S., Cornea. A. and Conn, P.M. (2009) Molecular mechanism of action of pharmacoperone rescue of misrouted GPCR mutants: The GnRH receptor. *Molecular Endocrinology*, **23**, 157-168.

doi:/10.1210/me.2008-0384

[58] Janovick, J.A., Maya-Nunez, G. and Conn, P.M. (2002) Rescue of hypogonadotropic hypogonadism-causing and manufactured GnRH receptor mutants by a specific protein-folding template: misrouted proteins as a novel disease etiology and therapeutic target. *The Journal of Clinical Endocrinology & Metabolism*, **87**, 3255-3262. doi:/10.1210/jc.87.7.3255

[59] Galietta, L.J., Springsteel, M.F., Eda, M., Niedzinski, E.J., By, K., Haddadin, M.J., Kurth, M.J., Nantz, M.H. and Verkman, A.S. (2001) Novel CFTR chloride channel activators identified by screening of combinatorial libraries based on flavone and benzoquinolizinium lead compounds. *The Journal of Biological Chemistry*, **276**, 19723-19728. doi:/10.1074/jbc.M101892200

[60] Ulloa-Aguirre, A., Janovick, J.A., Leanos-Miranda, A. and Conn, P.M. (2003) Misrouted cell surface receptors as a novel disease aetiology and potential therapeutic target: The case of hypogonadotropic hypogonadism due to gonadotropin-releasing hormone resistance. *Expert Opinion on Therapeutic Targets*, **7**, 175-185. doi:/10.1517/14728222.7.2.175

[61] Bernier, V., Lagace, M., Bichet, D.G. and Bouvier, M. (2004) Pharmacological chaperones: Potential treatment for conformational diseases. *Trends in Endocrinology & Metabolism*, **15**, 222-228. doi:/10.1016/j.tem.2004.05.003

[62] Noorwez, S.M., Malhotra, R., McDowell, J.H., Smith, K.A., Krebs, M.P. and Kaushal, S. (2004) Retinoids assist the cellular folding of the autosomal dominant retinitis pigmentosa opsin mutant P23H. *The Journal of Biological Chemistry*, **279**, 16278-16284. doi:/10.1074/jbc.M312101200

[63] Tveten, K., Holla, Ø.L., Ranheim, T., Berge, K.E., Leren, T.P. and Kulseth, M.A. (2007) 4-Phenylbutyrate restores the functionality of a misfolded mutant low-density lipoprotein receptor. *FEBS Journal*, **274**, 1881-1893. doi:/10.1111/j.1742-4658.2007.05735.x

[64] Benedek, G.B., Pande, J., Thurston, G.M. and Clark, J.I. (1999) Theoretical and experimental basis for the inhibition of cataract. *Progress in Retinal and Eye Research*, **18**, 391-402. doi:/10.1016/S1350-9462(98)00023-8

[65] Heiser, V., Scherzinger, E., Boeddrich, A., Nordhoff, E., Lurz, R., Schugardt, N., Lehrach, H. and Wanker, E.E. (2000) Inhibition of huntingtin fibrillogenesis by specific antibodies and small molecules: Implications for Huntington's disease therapy. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6739-6744. doi:/10.1073/pnas.110138997

[66] Muchowski, P.J. and Wacker, J.L. (2005) Modulation of neurodegeneration by molecular chaperones. *Nature Reviews Neuroscience*, **6**, 11-22. doi:/10.1038/nrn1587

[67] Forloni, G., Terreni, L., Bertani, I., Fogliarino, S., Invernizzi, R., Assini, A., Ribizzi, G., Negro, A., Calabrese, E., Volonté, M.A., Mariani, C., Franceschi, M., Tabaton, M. and Bertoli, A. (2002) Protein misfolding in Alzheimer's and Parkinson's disease: Genetics and molecular mechanisms. *Neurobiology of Aging*, **23**, 957-976. doi:/10.1016/S0197-4580(02)00076-3

[68] Peng, Y., Li, C., Chen, L., Sebti, S. and Chen, J. (2003) Rescue of mutant p53 transcription function by ellipticine. *Oncogene*, **22**, 4478-4487. doi:/10.1038/sj.onc.1206777

[69] Janovick, J.A., Goulet, M., Bush, E., Greer, J., Wettlaufer, D.G. and Conn, P.M. (2003) Structure-activity relations of successful pharmacologic chaperones for rescue of naturally occurring and manufactured mutants of the gonadotropin-releasing hormone receptor. *Journal of Pharmacology and Experimental Therapeutics*, **305**, 608- 614. doi:/10.1124/jpet.102.048454

[70] Costanzi, S. (2010) Modeling G protein-coupled receptors: A concrete possibility. *Chimica oggi*, **28**, 26-31.

[71] Sugahara, D., Kaji, H., Sugihara, K., Asano, M. and Narimatsu, H. (2012) Large-scale identification of target proteins of a glycosyltransferase isozyme by Lectin-IGOT-LC/MS, an LC/MS-based glycoproteomic approach. *Scientific Reports*, **2**, 680. doi:/10.1038/srep00680

[72] Ying, S.Y., Chang, D.C. and Lin, S.L. (2013) The MicroRNA. *Methods in Molecular Biology*, **936**, 1-19. doi:/10.1007/978-1-62703-083-0_1

[73] Coskun, M., Bjerrum, J.T., Seidelin, J.B. and Nielsen, O.H. (2012) MicroRNAs in inflammatory bowel disease—pathogenesis, diagnostics and therapeutics. *World Journal of Gastroenterology*, **18**, 4629-4634. doi:/10.3748/wjg.v18.i34.4629

Appendix: Algorithm and VBA Source Code of Scoring Matrix

```
Dim n As Integer, i As Integer, j As Integer, l As Integer,
na As Integer, nt As Integer, nc As Integer, ng As Integer
n = 1
Do Until Cells(n, 1) = ""
    n = n + 1
Loop
n = n - 1
For i = 1 To n - 1
    For j = i + 1 To n
        If Len(Cells(i, 1)) <> Len(Cells(j, 1)) Then
            MsgBox "Your sequences are not entered
with equal length!  "
            Exit For
        End If
    Next j
    If j <> n Then
        Exit For
    End If
    If Len(Cells(n, 1)) <> Len(Cells(1, 1)) Then
        Exit For
    End If
Next i
If i <> n - 1 And j <> n Then
    Cells(n + 2, 2) = "A"
    Cells(n + 3, 2) = "T"
    Cells(n + 4, 2) = "C"
    Cells(n + 5, 2) = "G"
    l = Len(Cells(1, 1))
    For i = 3 To l + 2
        Cells(n + 1, i) = i - 2
        Cells(n + 1, i). NumberFormatLocal = "0"
    Next i
    For i = 1 To l
        For j = 1 To n
            If Asc(Mid(Cells(j, 1), i, 1)) = 97 Then
                na = na + 1
            End If
            If Asc(Mid(Cells(j, 1), i, 1)) = 65 Then
                na = na + 1
            End If
            If Asc(Mid(Cells(j, 1), i, 1)) = 116 Then
                nt = nt + 1
            End If
            If Asc(Mid(Cells(j, 1), i, 1)) = 84 Then
                nt = nt + 1
            End If
            If Asc(Mid(Cells(j, 1), i, 1)) = 99 Then
                nc = nc + 1
            End If
            If Asc(Mid(Cells(j, 1), i, 1)) = 67 Then
                nc = nc + 1
            End If
            If Asc(Mid(Cells(j, 1), i, 1)) = 103 Then
                ng = ng + 1
            End If
            If Asc(Mid(Cells(j, 1), i, 1)) = 71 Then
                ng = ng + 1
            End If
        Next j
        Cells(n + 2, i + 2) = na / (n * l)
        Cells(n + 3, i + 2) = nt / (n * l)
        Cells(n + 4, i + 2) = nc / (n * l)
        Cells(n + 5, i + 2) = ng / (n * l)
        Cells(n + 2, i + 2). NumberFormatLocal =
"0.00"
        Cells(n + 3, i + 2).NumberFormatLocal =
"0.00"
        Cells(n + 4, i + 2). NumberFormatLocal =
"0.00"
        Cells(n + 5, i + 2). NumberFormatLocal =
"0.00"
        na = 0
        nt = 0
        nc = 0
        ng = 0
    Next i
    Cells(1, 1).Select
    End If
End Sub
```

Appendix 1. G-protein couple receptors (GPCRs) of GPCRs' database consisting of training set/test sets.

Swiss-Prot ID Homo sapiens	(AC)	Gene name	Chrom.	Locus	subfamily	Group*
AA1R_HUMAN	(P30542)	ADORA1	1	1q32.1	Adenosine type 1	A2, A14
AA3R_HUMAN	(P33765)	ADORA3	1	1p21 - p13	Adenosine type 3	T23
CELR2_HUMAN	(Q9HCU4)	CELSR2	1	1p21	Cadherin EGF LAG (CELSR)	B15
CNR2_HUMAN	(P34972)	CNR2	1	1p36.11	Cannabinoid	A1, A14
GNRR2_HUMAN	(Q96P88)	GNRHR2	1	1q12	Gonadotropin-releasing hormone type II	A1
LGR6_HUMAN	(Q9HBX8)	LGR6	1	1q32.1	LGR like (hormone receptors)	A1, A14
ACM3_HUMAN	(P20309)	CHRM3	1	1q41 - q44	Musc. acetylcholine Vertebrate type 3	A2
OR1C1_HUMAN	(Q15619)	OR1C1	1	1q44	Olfactory II fam 1/MOR125 - 138, 156	O6, O17
O10K1_HUMAN	(Q8NGX5)	OR10K1	1	1q23.1	Olfactory II fam 10/MOR263 - 269	O7
O10K2_HUMAN	(Q6IF99)	OR10K2	1	1q23.1	Olfactory II fam 10/MOR263 - 269	O7, O17
O10R2_HUMAN	(Q8NGX6)	OR10R2	1	1q23.1	Olfactory II fam 10/MOR263 - 269	O7, O17
O10T2_HUMAN	(Q8NGX3)	OR10T2	1	1q23.1	Olfactory II fam 10/MOR263 - 269	O7
O10X1_HUMAN	(Q8NGY0)	OR10X1	1	1q23.1	Olfactory II fam 10/MOR263 - 269	O7
O10Z1_HUMAN	(Q8NGY1)	OR10Z1	1	1q23.1	Olfactory II fam 10/MOR263 - 269	O7
O10J1_HUMAN	(P30954)	OR10J1	1	1q23.2	Olfactory II fam 10/MOR263 - 269	O7
O10J3_HUMAN	(Q5JRS4)	OR10J3	1	1q23.2	Olfactory II fam 10/MOR263 - 269	O7
O10J5_HUMAN	(Q8NHC4)	OR10J5	1	1q23.2	Olfactory II fam 10/MOR263 - 269	O7
O11L1_HUMAN	(Q8NGX0)	OR11L1	1	1q44	Olfactory II fam 11/MOR106, 121 - 122	O17
O13G1_HUMAN	(Q8NGZ3)	OR13G1	1	1q44	Olfactory II fam 13/MOR253	O8, O17
O2T10_HUMAN	(Q8NGZ9)	OR2T10	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9, O17
O2T11_HUMAN	(Q8NH01)	OR2T11	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9, O17
O2T12_HUMAN	(Q8NG77)	OR2T12	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9, O17
O2T27_HUMAN	(Q8NH04)	OR2T27	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9, D18
O2T33_HUMAN	(Q8NG76)	OR2T33	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
O2T34_HUMAN	(Q8NGX1)	OR2T34	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
O2T35_HUMAN	(Q8NGX2)	OR2T35	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2BB_HUMAN	(Q5JQS5)	OR2B11	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2C3_HUMAN	(Q8N628)	OR2C3	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2G3_HUMAN	(Q8NGZ4)	OR2G3	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2G6_HUMAN	(Q5TZ20)	OR2G6	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2L2_HUMAN	(Q8NH16)	OR2L2	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2L3_HUMAN	(Q8NG85)	OR2L3	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2L8_HUMAN	(Q8NGY9)	OR2L8	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2LD_HUMAN	(Q8N349)	OR2L13	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2M2_HUMAN	(Q96R28)	OR2M2	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9

Continued

OR2M3_HUMAN	(Q8NG83)	OR2M3	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2M4_HUMAN	(Q96R27)	OR2M4	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2M7_HUMAN	(Q8NG81)	OR2M7	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2T2_HUMAN	(Q6IF00)	OR2T2	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2T3_HUMAN	(Q8NH03)	OR2T3	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2T5_HUMAN	(Q6IEZ7)	OR2T5	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2T6_HUMAN	(Q8NHC8)	OR2T6	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2W3_HUMAN	(Q7Z3T1)	OR2W3	1	1q44	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
O4F29_HUMAN	(Q6IEY1)	OR4F29	1	1p36.33	Olfactory II fam 4/MOR225 - 248	O10, O17
OR4F5_HUMAN	(Q8NH21)	OR4F5	1	1p36.33	Olfactory II fam 4/MOR225 - 248	O10, O17
O5AT1_HUMAN	(Q8NHC5)	OR5AT1	1	1q44	Olfactory II fam 5/MOR172 - 224, 249, 254	O11, O17
O5BF1_HUMAN	(Q8NHC7)	OR5BF1	1	1q44	Olfactory II fam 5/MOR172 - 224, 249, 254	O11, O17
OR6K2_HUMAN	(Q8NGY2)	OR6K2	1	1q23.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12, O17
OR6K6_HUMAN	(Q8NGW6)	OR6K6	1	1q23.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12, O17
OR6N1_HUMAN	(Q8NGY5)	OR6N1	1	1q23.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12, O17
OR6N2_HUMAN	(Q8NGY6)	OR6N2	1	1q23.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6Y1_HUMAN	(Q8NGX8)	OR6Y1	1	1q23.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6F1_HUMAN	(Q8NGZ6)	OR6F1	1	1q44	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OPRD_HUMAN	(P41143)	OPRD1	1	1p36.1 - p34.3	Opioid type D	A3, A14
OX1R_HUMAN	(O43613)	HCRTR1	1	1p33	Orexin	A1, A14
PTAFR_HUMAN	(P25105)	PTAFR	1	1p35 - p34.3	Platelet activating factor	A4, A14, D18
PF2R_HUMAN	(P43088)	PTGFR	1	1p31.1	Prostaglandin F2-alpha	T23
OPN3_HUMAN	(Q9H1Y3)	OPN3	1	1q43	Rhodopsin Other	A1, A14
5HT1D_HUMAN	(P28221)	HTR1D	1	1p36.3 - p34.3	Serotonin type 1	A2, A14
5HT6R_HUMAN	(P50406)	HTR6	1	1p36 - p35	Serotonin type 6	T23
EDG1_HUMAN	(P21453)	EDG1	1	1p21	Sphingosine 1-phosphate Edg-1	A1, A14
V1BR_HUMAN	(P47901)	AVPR1B	1	1q32	Vasopressin type 1	A1, A14
ADA2B_HUMAN	(P18089)	ADRA2B	2	2p13 - q13	Alpha Adrenoceptors type 2	A2
CALRL_HUMAN	(Q16602)	CALCRL	2	2q32.1	Calcitonin	B15, D18
FZD7_HUMAN	(O75084)	FZD7	2	2q33	frizzled Group A (Fz 1 & 2 & 4 & 5 & 7 - 9)	F22
FZD5_HUMAN	(Q13467)	FZD5	2	2q33 - q34	frizzled Group A (Fz 1 & 2 & 4 & 5 & 7 - 9)	F22, D18
NMUR1_HUMAN	(Q9HB89)	NMUR1	2	2q37.1	Neuromedin U	A2, A14
NTR2_HUMAN	(O95665)	NTSR2	2	2p25.1	Neurotensin	A2, A14
OR6B2_HUMAN	(Q6IFH4)	OR6B2	2	2q37.3	Olfactory II fam 6/MOR103 - 105, 107 - 119	A12
OR6B3_HUMAN	(Q8NGW1)	OR6B3	2	2q37.3	Olfactory II fam 6/MOR103 - 105, 107 - 119	A12
PTHR2_HUMAN	(P49190)	PTHR2	2	2q33	Parathyroid hormone	B15
GPR35_HUMAN	(Q9HC97)	GPR35	2	2q37.3	Purinoceptor P2RY5, 8, 9, 10 GPR35, 92, 174	D18

Continued

5HT2B_HUMAN	(P41595)	HTR2B	2	2q36.3 - q37.1	Serotonin type 2	A2, A14
AGTR1_HUMAN	(P30556)	AGTR1	3	3q21 - q25	Angiotensin type 1	S19
CELR3_HUMAN	(Q9NYQ7)	CELSR3	3	3p24.1 - p21.2	Cadherin EGF LAG (CELSR)	S19
O75307_HUMAN	(O75307)	CCRL2	3	3p21	C-C Chemokine other	3, 14
CCR1_HUMAN	(P32246)	CCR1	3	3p21	C-C Chemokine type 1	A3, S19
CCRL1_HUMAN	(Q9NPB9)	CCRL1	3	3q22	C-C Chemokine type 11	A3, S19
CCR2_HUMAN	(P41597)	CCR2	3	3p21	C-C Chemokine type 2	A3, S19
CCR3_HUMAN	(P51677)	CCR3	3	3p21.3	C-C Chemokine type 3	A3, S19
CCR4_HUMAN	(P51679)	CCR4	3	3p24	C-C Chemokine type 4	A3, S19
CCR5_HUMAN	(P51681)	CCR5	3	3p21	C-C Chemokine type 5	A3, S19
CCR8_HUMAN	(P51685)	CCR8	3	3p22	C-C Chemokine type 8	A3, S19
CCR9_HUMAN	(P51686)	CCR9	3	3p21.3	C-C Chemokine type 9	A3, S19
CCBP2_HUMAN	(O00590)	CCBP2	3	3p21.3	C-C Chemokine type X	A3, S19
CX3C1_HUMAN	(P49238)	CX3CR1	3	3p21|3p21.3	C-X3-C Chemokine	S19
CXCR6_HUMAN	(O00574)	CXCR6	3	3p21	C-X-C Chemokine type 6 (Bonzo)	T23, T24
DRD3_HUMAN	(P35462)	DRD3	3	3q13.3	Dopamine Vertebrate type 3	A2, D18, S19
GP156_HUMAN	(Q8NFN8)	GPR156	3	3q13.33	GABA-B like	S19
GPR15_HUMAN	(P49685)	GPR15	3	3q11.2 - q13.1	GPR	S19
HRH1_HUMAN	(P35367)	HRH1	3	3p25	Histamine type 1	A2, A14, S19
MGR2_HUMAN	(Q14416)	GRM2	3	3p21.2	Metabotropic glutamate group II	C16
MGR7_HUMAN	(Q14831)	GRM7	3	3p26.1 - p25.1	Metabotropic glutamate group III	C16, S19
O5AC2_HUMAN	(Q9NZP5)	OR5AC2	3	3q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11, O17, S19
OR5H2_HUMAN	(Q8NGV7)	OR5H2	3	3q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5H6_HUMAN	(Q8NGV6)	OR5H6	3	3q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5K1_HUMAN	(Q8NHB7)	OR5K1	3	3q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5K2_HUMAN	(Q8NHB8)	OR5K2	3	3q12.2	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OXYR_HUMAN	(P30559)	OXTR	3	3p25	Oxytocin/mesotocin	S19
PTHR1_HUMAN	(Q03431)	PTHR1	3	3p22 - p21.1	Parathyroid hormone	B15, S19
P2Y14_HUMAN	(Q15391)	P2RY14	3	3q21 - q25	Purinoceptor P2RY12-14 GPR87 (UDP-Glucose)	A4, A14, S19
GPR87_HUMAN	(Q9BY21)	GPR87	3	3q24	Purinoceptor P2RY12-14 GPR87 (UDP-Glucose)	A4
P2Y13_HUMAN	(Q9BPV8)	P2RY13	3	3q24	Purinoceptor P2RY12-14 GPR87 (UDP-Glucose)	A4
P2Y12_HUMAN	(Q9H244)	P2RY12	3	3q24 - q25	Purinoceptor P2RY12-14 GPR87 (UDP-Glucose)	T23, T24
P2RY1_HUMAN	(P47900)	P2RY1	3	3q25.2	Purinoceptor P2RY1-4, 6, 11 GPR91	A4
GPR62_HUMAN	(Q9BZJ7)	GPR62	3	3p21.1	Putative/unclassified Class A GPCRs	S19
GP128_HUMAN	(Q96K78)	GPR128	3	3q12.2	Putative/unclassified Class B GPCRs	S19
GP175_HUMAN	(Q86W33)	GPR175	3	3q21.2	Putative/unclassified other	S19
OPSD_HUMAN	(P08100)	RHO	3	3q21 - q24	Rhodopsin Vertebrate type 1	A1, S19

Continued

5HT1F_HUMAN	(P30939)	HTR1F	3	3p12	Serotonin type 1	A2, S19
GPR27_HUMAN	(Q9NS67)	GPR27	3	3p21 - p14	SREB	A1, A14, S19
ADA2C_HUMAN	(P18825)	ADRA2C	4	4p16	Alpha Adrenoceptors type 2	T23, T24
CCKAR_HUMAN	(P32238)	CCKAR	4	4p15.1 - p15.2	CCK type A	A1, A14
DRD5_HUMAN	(P21918)	DRD5	4	4p16.1	Dopamine Vertebrate type 1	A2, A14
EDNRA_HUMAN	(P25101)	EDNRA	4	4q31.22 - q31.23	Endothelin	A1, A14
GNRHR_HUMAN	(P30968)	GNRHR	4	4q21.2	Gonadotropin-releasing hormone type I	A14, D18
Q2M215_HUMAN	(Q2M215)	RXFP1	4	4q32.1	LGR like (hormone receptors)	A1, D18
MTR1A_HUMAN	(P48039)	MTNR1A	4	4q35.1	Melatonin	A1, A14
NPFF2_HUMAN	(Q9Y5X5)	NPFFR2	4	4q21	Neuropeptide FF	A1, A14
NPY1R_HUMAN	(P25929)	NPY1R	4	4q31.3 - q32	Neuropeptide Y type 1	A1, D18
NPY2R_HUMAN	(P49146)	NPY2R	4	4q31	Neuropeptide Y type 2	A1, D21
NPY5R_HUMAN	(Q15761)	NPY5R	4	4q31 - q32	Neuropeptide Y type 5	T23
OPSX_HUMAN	(O14718)	RRH	4	4q25	Rhodopsin Other	A1, D18
ADA1B_HUMAN	(P35368)	ADRA1B	5	5q23 - q32	Alpha Adrenoceptors type 1	A2, A14
ADRB2_HUMAN	(P07550)	ADRB2	5	5q31 - q32	Beta Adrenoceptors type 2	A2, D18
DRD1_HUMAN	(P21728)	DRD1	5	5q35.1	Dopamine Vertebrate type 1	A2
HRH2_HUMAN	(P25021)	HRH2	5	5q35.2	Histamine type 2	A2
MGR6_HUMAN	(O15303)	GRM6	5	5q35	Metabotropic glutamate group III	C16
NMUR2_HUMAN	(Q9GZQ4)	NMUR2	5	5q33.1	Neuromedin U	A2
OR2V2_HUMAN	(Q96R30)	OR2V2	5	5q35.3	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2Y1_HUMAN	(Q8NGV0)	OR2Y1	5	5q35.3	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR4F3_HUMAN	(O95013)	OR4F3	5	5q35.3	Olfactory II fam 4/MOR225 - 248	O10, O17
PE2R4_HUMAN	(P35408)	PTGER4	5	5p13.1	Prostaglandin E2 subtype EP4	A1
PAR2_HUMAN	(P55085)	F2RL1	5	5q13	Proteinase-activated	A4, A14
PAR3_HUMAN	(O00254)	F2RL2	5	5q13	Proteinase-activated	A4, D18
5HT1A_HUMAN	(P08908)	HTR1A	5	5q11.2 - q13	Serotonin type 1	A2
5HT4R_HUMAN	(Q13639)	HTR4	5	5q31 - q33	Serotonin type 4	A2
TA2R1_HUMAN	(Q9NYW7)	TAS2R1	5	5p15	Taste receptors T2R	D18
NMBR_HUMAN	(P28336)	NMBR	6	6q21 - qter	Bombesin	A1, A14
BAI3_HUMAN	(O60242)	BAI3	6	6q12	Brain-specific angiogenesis inhibitor (BAI)	B15
CNR1_HUMAN	(P21554)	CNR1	6	6q14 - q15	Cannabinoid	A1
O00421_HUMAN	(O00421)	ccr6	6	6q27	C-C Chemokine other	A3, D18
GABR1_HUMAN	(Q9UBS5)	GABBR1	6	6p21.31	GABA-B subtype 1	C16
GLP1R_HUMAN	(P43220)	GLP1R	6	6p21	Glucagon	B15
MAS1L_HUMAN	(P35410)	MAS1L	6	6p21	Mas proto-oncogene & Mas-related (MRGs)	A4, A14
MAS_HUMAN	(P04201)	MAS1	6	6q25.3 - q26	Mas proto-oncogene & Mas-related (MRGs)	A4

Continued

MGR1_HUMAN	(Q13255)	GRM1	6	6q24	Metabotropic glutamate group I	C16
MGR4_HUMAN	(Q14833)	GRM4	6	6p21.3	Metabotropic glutamate group III	C16
O10C1_HUMAN	(Q96KK4)	OR10C1	6	6p22.1	Olfactory II fam 10/MOR263 - 269	O7
O12D3_HUMAN	(Q9UGF7)	OR12D3	6	6p22.1	Olfactory II fam 12/MOR250	O17
OR2B6_HUMAN	(P58173)	OR2B6	6	6p21.3	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2H1_HUMAN	(Q9GZK4)	OR2H1	6	6p21.3	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2H2_HUMAN	(O95918)	OR2H2	6	6p21.3	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2B2_HUMAN	(Q9GZK3)	OR2B2	6	6p21.3 - 22.3	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2W1_HUMAN	(Q9Y3N9)	OR2W1	6	6p21.31 - 21.33	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2J2_HUMAN	(O76002)	OR2J2	6	6p21.31 - 22.2	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2B3_HUMAN	(O76000)	OR2B3	6	6p22.1	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2J3_HUMAN	(O76001)	OR2J3	6	6p22.1	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2A4_HUMAN	(O95047)	OR2A4	6	6q23	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR5U1_HUMAN	(Q9UGF5)	OR5U1	6	6p22.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5V1_HUMAN	(Q9UGF6)	OR5V1	6	6p22.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OPRM_HUMAN	(P35372)	OPRM1	6	6q24 - q25	Opioid type M	A3, D18
OX2R_HUMAN	(O43614)	HCRTR2	6	6p11 - q11	Orexin	A1, D18
OPN5_HUMAN	(Q6U736)	OPN5	6	6p12.3	Rhodopsin Other	A1
5HT1B_HUMAN	(P28222)	HTR1B	6	6q13	Serotonin type 1	A2
5HT1E_HUMAN	(P28566)	HTR1E	6	6q14 - q15	Serotonin type 1	T23, T24
Q2M1V1_HUMAN	(Q2M1V1)	TAAR5	6	6q23	Trace amine	A2, A14
Q2M1W5_HUMAN	(Q2M1W5)	TAAR1	6	6q23.2	Trace amine	A2, D18
TAAR6_HUMAN	(Q96RI8)	TAAR6	6	6q23.2	Trace amine	A2
TAAR8_HUMAN	(Q969N4)	TAAR8	6	6q23.2	Trace amine	A2
TAAR9_HUMAN	(Q96RI9)	TAAR9	6	6q23.2	Trace amine	A2
TAAR2_HUMAN	(Q9P1P5)	TAAR2	6	6q24	Trace amine	T23, T24
CML2_HUMAN	(Q99527)	GPR30	7	7p22	Chemokine receptor-like 2	T23
CRFR2_HUMAN	(Q13324)	CRHR2	7	7p14.3	Corticotropin releasing factor	B15
FZD9_HUMAN	(O00144)	FZD9	7	7q11.23	frizzled Group A (Fz 1 & 2 & 4 & 5 & 7 - 9)	F22
FZD1_HUMAN	(Q9UP38)	FZD1	7	7q21	frizzled Group A (Fz 1 & 2 & 4 & 5 & 7 - 9)	F22
MGR3_HUMAN	(Q14832)	GRM3	7	7q21.1 - q21.2	Metabotropic glutamate group II	C16
MGR8_HUMAN	(O00222)	GRM8	7	7q31.3 - q32.1	Metabotropic glutamate group III	T23, T24
ACM2_HUMAN	(P08172)	CHRM2	7	7q31 - q35	Musc. acetylcholine Vertebrate type 2	A2, D18
O2AE1_HUMAN	(Q8NHA4)	OR2AE1	7	7q22.1	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
O2A12_HUMAN	(Q8NGT7)	OR2A12	7	7q35	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
O2A42_HUMAN	(Q8NGT9)	OR2A42	7	7q35	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2A2_HUMAN	(Q6IF42)	OR2A2	7	7q35	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9

Continued

OR2A5_HUMAN	(Q96R48)	OR2A5	7	7q35	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2A7_HUMAN	(Q96R45)	OR2A7	7	7q35	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2F1_HUMAN	(Q13607)	OR2F1	7	7q35	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2F2_HUMAN	(O95006)	OR2F2	7	7q35	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR6B1_HUMAN	(O95007)	OR6B1	7	7q35	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6V1_HUMAN	(Q8N148)	OR6V1	7	7q35	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR9A2_HUMAN	(Q8NGT5)	OR9A2	7	7q34	Olfactory II fam 9/MOR120	O12, O17
OPSB_HUMAN	(P03999)	OPN1SW	7	7q31.3 - q32	Rhodopsin Vertebrate type 3	T23
5HT5A_HUMAN	(P47898)	HTR5A	7	7q36.1	Serotonin type 5	A2
GPR85_HUMAN	(P60893)	GPR85	7	7q31	SREB	A1, D18
VIPR2_HUMAN	(P41587)	VIPR2	7	7q36.3	Vasoactive intestinal polypeptide	B15, D18
ADA1A_HUMAN	(P35348)	ADRA1A	8	8p21 - p11.2	Alpha Adrenoceptors type 1	A2, D18
ADRB3_HUMAN	(P13945)	ADRB3	8	8p12 - p11.2	Beta Adrenoceptors type 3	T23
BAI1_HUMAN	(O14514)	BAI1	8	8q24	Brain-specific angiogenesis inhibitor (BAI)	B15
FZD3_HUMAN	(Q9NPG1)	FZD3	8	8p21	frizzled Group B (Fz 3 & 6)	F22
FZD6_HUMAN	(O60353)	FZD6	8	8q22.3 - q23.1	frizzled Group B (Fz 3 & 6)	T23, T24
OR1B1_HUMAN	(Q8NGR6)	OR1B1	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6, O17
OR1J1_HUMAN	(Q8NGS3)	OR1J1	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6, O17
OR1K1_HUMAN	(Q8NGR3)	OR1K1	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6, O17
OR1L3_HUMAN	(Q8NH93)	OR1L3	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1L4_HUMAN	(Q8NGR5)	OR1L4	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1L8_HUMAN	(Q8NGR8)	OR1L8	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1N1_HUMAN	(Q8NGS0)	OR1N1	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1N2_HUMAN	(Q8NGR9)	OR1N2	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1Q1_HUMAN	(Q15612)	OR1Q1	9	9q33.2	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1J2_HUMAN	(Q8NGS2)	OR1J2	9	9q34.11	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1J4_HUMAN	(Q8NGS1)	OR1J4	9	9q34.11	Olfactory II fam 1/MOR125 - 138, 156	O6
OR2S1_HUMAN	(Q9NQN1)	OR2S2	9	9p13.1 - 13.3	Olfactory II fam 13/MOR253	O8, O17
O13J1_HUMAN	(Q8NGT2)	OR13J1	9	9p13.3	Olfactory II fam 13/MOR253	O8, O17
O13C2_HUMAN	(Q8NGS9)	OR13C2	9	9q31.1	Olfactory II fam 13/MOR253	O8
O13C3_HUMAN	(Q8NGS6)	OR13C3	9	9q31.1	Olfactory II fam 13/MOR253	O8
O13C4_HUMAN	(Q8NGS5)	OR13C4	9	9q31.1	Olfactory II fam 13/MOR253	O8
O13C5_HUMAN	(Q8NGS8)	OR13C5	9	9q31.1	Olfactory II fam 13/MOR253	O8
O13C8_HUMAN	(Q8NGS7)	OR13C8	9	9q31.1	Olfactory II fam 13/MOR253	O8
O13C9_HUMAN	(Q8NGT0)	OR13C9	9	9q31.1	Olfactory II fam 13/MOR253	O8
O13F1_HUMAN	(Q8NGS4)	OR13F1	9	9q31.1	Olfactory II fam 13/MOR253	O8
OR5C1_HUMAN	(Q8NGR4)	OR5C1	9	9q33.2	Olfactory II fam 5/MOR172 - 224, 249, 254	O11

Continued

Q5VTM0_HUMAN	(Q5VTM0)	SLC31A2	9	9q31 - q32	Orexin	T23, T24
EDG3_HUMAN	(Q99500)	EDG3	9	9q22.1 - q22.2	Sphingosine 1-phosphate Edg-3	A1, D18
ADRB1_HUMAN	(P08588)	ADRB1	10	10q24 - q26	Beta Adrenoceptors type 1	A2, A14
FZD8_HUMAN	(Q9H461)	FZD8	10	10p11.21	frizzled Group A (Fz 1 & 2 & 4 & 5 & 7 - 9)	F22
NPFF1_HUMAN	(Q9GZQ6)	NPFFR1	10	10q21 - q22	Neuropeptide FF	A1
NPY4R_HUMAN	(P50391)	PPYR1	10	10q11.2	Neuropeptide Y type 4	A1
GP158_HUMAN	(Q5T848)	GPR158	10	10p12.1	Putative/unclassified Class C GPCRs	T23
RGR_HUMAN	(P47804)	RGR	10	10q23	Rhodopsin Other	A1
NK2R_HUMAN	(P21452)	TACR2	10	10q11 - q21	Substance K (NK2)	D21
APJ_HUMAN	(P35414)	AGTRL1	11	11q12	APJ like	S20
GASR_HUMAN	(P32239)	CCKBR	11	11p15.4	CCK type B	A1, S20
GPR44_HUMAN	(Q9Y5Y4)	GPR44	11	11q12 - q13.3	Chemokine receptor-like 1	D18, S20
CXCR5_HUMAN	(P32302)	BLR1	11	11q23.3	C-X-C Chemokine type 5	A3
DRD2_HUMAN	(P14416)	DRD2	11	11q23	Dopamine Vertebrate type 2	A2, A14, S20
FZD4_HUMAN	(Q9ULV1)	FZD4	11	11q14.2	frizzled Group A (Fz 1&2&4&5&7-9)	F22, S20
LGR4_HUMAN	(Q9BXB1)	LGR4	11	11p14 - p13	LGR like (hormone receptors)	A1, S20
MRGX2_HUMAN	(Q96LB1)	MRGPRX2	11	11p15.1	Mas proto-oncogene & Mas-related (MRGs)	A4, S20
MRGX3_HUMAN	(Q96LB0)	MRGPRX3	11	11p15.1	Mas proto-oncogene & Mas-related (MRGs)	A4
MRGX4_HUMAN	(Q96LA9)	MRGPRX4	11	11p15.1	Mas proto-oncogene & Mas-related (MRGs)	A4
MRGX1_HUMAN	(Q96LB2)	MRGPRX1	11	11p15.1|11	Mas proto-oncogene & Mas-related (MRGs)	A4
Q2M1V7_HUMAN	(Q2M1V7)	MRGPRE	11	11p15.4	Mas proto-oncogene & Mas-related (MRGs)	A4
MRGRD_HUMAN	(Q8TDS7)	MRGPRD	11	11q13.3	Mas proto-oncogene & Mas-related (MRGs)	T23
MTR1B_HUMAN	(P49286)	MTNR1B	11	11q21 - q22	Melatonin	A1, S20
MGR5_HUMAN	(P41594)	GRM5	11	11q14.3	Metabotropic glutamate group I	C16, S20
ACM1_HUMAN	(P11229)	CHRM1	11	11q13	Musc. acetylcholine Vertebrate type 1	A2, A14, S20
ACM4_HUMAN	(P08173)	CHRM4	11	11p12 - p11.2	Musc. acetylcholine Vertebrate type 4	A2
GPR83_HUMAN	(Q9NYM4)	GPR83	11	11q21	Neuropeptide Y other	A1, A14, S20
O51B2_HUMAN	(Q9Y5P1)	OR51B2	11	11p15	Olfactory I fam 51 - 52/MOR1 - 42	O5, O17, S20
O51B4_HUMAN	(Q9Y5P0)	OR51B4	11	11p15	Olfactory I fam 51 - 52/MOR1 - 42	O5, O17
O51E2_HUMAN	(Q9H255)	OR51E2	11	11p15	Olfactory I fam 51 - 52/MOR1 - 42	O5, O17
O51A2_HUMAN	(Q8NGJ7)	OR51A2	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5, D18
O51A4_HUMAN	(Q8NGJ6)	OR51A4	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51A7_HUMAN	(Q8NH64)	OR51A7	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51B5_HUMAN	(Q9H339)	OR51B5	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51B6_HUMAN	(Q9H340)	OR51B6	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51D1_HUMAN	(Q8NGF3)	OR51D1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51G1_HUMAN	(Q8NGK1)	OR51G1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5

Continued

O51G2_HUMAN	(Q8NGK0)	OR51G2	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51I1_HUMAN	(Q9H343)	OR51I1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51I2_HUMAN	(Q9H344)	OR51I2	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51L1_HUMAN	(Q8NGJ5)	OR51L1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51Q1_HUMAN	(Q8NH59)	OR51Q1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51S1_HUMAN	(Q8NGJ8)	OR51S1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O51V1_HUMAN	(Q9H2C8)	OR51V1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52A5_HUMAN	(Q9H2C5)	OR52A5	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52B2_HUMAN	(Q96RD2)	OR52B2	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52B4_HUMAN	(Q8NGK2)	OR52B4	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52B6_HUMAN	(Q8NGF0)	OR52B6	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52D1_HUMAN	(Q9H346)	OR52D1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52E2_HUMAN	(Q8NGJ4)	OR52E2	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52E4_HUMAN	(Q8NGH9)	OR52E4	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52E6_HUMAN	(Q96RD3)	OR52E6	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52H1_HUMAN	(Q8NGJ2)	OR52H1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52I1_HUMAN	(Q8NGK6)	OR52I1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52I2_HUMAN	(Q8NH67)	OR52I2	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52J3_HUMAN	(Q8NH60)	OR52J3	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52K1_HUMAN	(Q8NGK4)	OR52K1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52K2_HUMAN	(Q8NGK3)	OR52K2	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52M1_HUMAN	(Q8NGK5)	OR52M1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52N1_HUMAN	(Q8NH53)	OR52N1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52N2_HUMAN	(Q8NGI0)	OR52N2	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52N4_HUMAN	(Q8NGI2)	OR52N4	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52N5_HUMAN	(Q8NH56)	OR52N5	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52W1_HUMAN	(Q6IF63)	OR52W1	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O56A3_HUMAN	(Q8NH54)	OR56A3	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O56B4_HUMAN	(Q8NH76)	OR56B4	11	11p15.4	Olfactory I fam 51 - 52/MOR1 - 42	O5
O52A1_HUMAN	(Q9UKL2)	OR52A1	11	11p15.5	Olfactory I fam 51 - 52/MOR1 - 42	T23
OR1S1_HUMAN	(Q8NH92)	OR1S1	11	11q12.1	Olfactory II fam 1/MOR125 - 138, 156	O6, S20
OR1S2_HUMAN	(Q8NGQ3)	OR1S2	11	11q12.1	Olfactory II fam 1/MOR125 - 138, 156	O6
O10A3_HUMAN	(P58181)	OR10A3	11	11p15.4	Olfactory II fam 10/MOR263 - 269	O7, S20
O10A4_HUMAN	(Q9H209)	OR10A4	11	11p15.4	Olfactory II fam 10/MOR263 - 269	O7
O10A5_HUMAN	(Q9H207)	OR10A5	11	11p15.4	Olfactory II fam 10/MOR263 - 269	O7
O10A6_HUMAN	(Q8NH74)	OR10A6	11	11p15.4	Olfactory II fam 10/MOR263 - 269	O7
O10Q1_HUMAN	(Q8NGQ4)	OR10Q1	11	11q12.1	Olfactory II fam 10/MOR263 - 269	O7

Continued

O10W1_HUMAN	(Q8NGF6)	OR10W1	11	11q12.1	Olfactory II fam 10/MOR263 - 269	O7
O10AG_HUMAN	(Q8NH19)	OR10AG1	11	11q12.2	Olfactory II fam 10/MOR263 - 269	O7
O10V1_HUMAN	(Q8NGI7)	OR10V1	11	11q12.2	Olfactory II fam 10/MOR263 - 269	O7
O10G4_HUMAN	(Q8NGN3)	OR10G4	11	11q24.1	Olfactory II fam 10/MOR263 - 269	O7
O10G7_HUMAN	(Q8NGN6)	OR10G7	11	11q24.1	Olfactory II fam 10/MOR263 - 269	O7
O10G8_HUMAN	(Q8NGN5)	OR10G8	11	11q24.1	Olfactory II fam 10/MOR263 - 269	O7
O10G9_HUMAN	(Q8NGN4)	OR10G9	11	11q24.1	Olfactory II fam 10/MOR263 - 269	O7
O10S1_HUMAN	(Q8NGN2)	OR10S1	11	11q24.1	Olfactory II fam 10/MOR263 - 269	O7
O2AG1_HUMAN	(Q9H205)	OR2AG1	11	11p15.4	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9, S20
OR2D2_HUMAN	(Q9H210)	OR2D2	11	11p15.4	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR2D3_HUMAN	(Q8NGH3)	OR2D3	11	11p15.4	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR4A5_HUMAN	(Q8NH83)	OR4A5	11	11p11.12	Olfactory II fam 4/MOR225 - 248	O10, D18, S20
OR4CC_HUMAN	(Q96R67)	OR4C12	11	11p11.12	Olfactory II fam 4/MOR225 - 248	O10
OR4CD_HUMAN	(Q8NGP0)	OR4C13	11	11p11.12	Olfactory II fam 4/MOR225 - 248	O10
O4A47_HUMAN	(Q6IF82)	OR4A47	11	11p11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4B1_HUMAN	(Q8NGF8)	OR4B1	11	11p11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4S1_HUMAN	(Q8NGB4)	OR4S1	11	11p11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4X1_HUMAN	(Q8NH49)	OR4X1	11	11p11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4X2_HUMAN	(Q8NGF9)	OR4X2	11	11p11.2	Olfactory II fam 4/MOR225 - 248	O10
O4A16_HUMAN	(Q8NH70)	OR4A16	11	11q11	Olfactory II fam 4/MOR225 - 248	O10
OR4C6_HUMAN	(Q8NH72)	OR4C6	11	11q11	Olfactory II fam 4/MOR225 - 248	O10
OR4CB_HUMAN	(Q6IEV9)	OR4C11	11	11q11	Olfactory II fam 4/MOR225 - 248	O10
OR4CG_HUMAN	(Q8NGL9)	OR4C16	11	11q11	Olfactory II fam 4/MOR225 - 248	O10
OR4S2_HUMAN	(Q8NH73)	OR4S2	11	11q11	Olfactory II fam 4/MOR225 - 248	O10
OR4D6_HUMAN	(Q8NGJ1)	OR4D6	11	11q12.1	Olfactory II fam 4/MOR225 - 248	O10
OR4D9_HUMAN	(Q8NGE8)	OR4D9	11	11q12.1	Olfactory II fam 4/MOR225 - 248	O10
OR4DA_HUMAN	(Q8NGI6)	OR4D10	11	11q12.1	Olfactory II fam 4/MOR225 - 248	O10
OR4DB_HUMAN	(Q8NGI4)	OR4D11	11	11q12.1	Olfactory II fam 4/MOR225 - 248	O10
OR4P4_HUMAN	(Q8NGL7)	OR4P4	11	11q12.1	Olfactory II fam 4/MOR225 - 248	O10
OR4D5_HUMAN	(Q8NGN0)	OR4D5	11	11q24.1	Olfactory II fam 4/MOR225 - 248	O10
OR5P2_HUMAN	(Q8WZ92)	OR5P2	11	11p15.4	Olfactory II fam 5/MOR172 - 224, 249, 254	O11, S20
OR5P3_HUMAN	(Q8WZ94)	OR5P3	11	11p15.4	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
O5AP2_HUMAN	(Q8NGF4)	OR5AP2	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
O5AR1_HUMAN	(Q8NGP9)	OR5AR1	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
O5AS1_HUMAN	(Q8N127)	OR5AS1	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5DD_HUMAN	(Q8NGL4)	OR5D13	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5DE_HUMAN	(Q8NGL3)	OR5D14	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11

Continued

OR5DI_HUMAN	(Q8NGL1)	OR5D18	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5F1_HUMAN	(O95221)	OR5F1	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5I1_HUMAN	(Q13606)	OR5I1	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5L1_HUMAN	(Q8NGL2)	OR5L1	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5L2_HUMAN	(Q8NGL0)	OR5L2	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5M1_HUMAN	(Q8NGP8)	OR5M1	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5M3_HUMAN	(Q8NGP4)	OR5M3	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5M9_HUMAN	(Q8NGP3)	OR5M9	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5MA_HUMAN	(Q6IEU7)	OR5M10	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5MB_HUMAN	(Q96RB7)	OR5M11	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5T1_HUMAN	(Q8NG75)	OR5T1	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5T2_HUMAN	(Q8NGG2)	OR5T2	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5T3_HUMAN	(Q8NGG3)	OR5T3	11	11q11	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
O5AK2_HUMAN	(Q8NH90)	OR5AK2	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
O5AN1_HUMAN	(Q8NGI8)	OR5AN1	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5A1_HUMAN	(Q8NGJ0)	OR5A1	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5A2_HUMAN	(Q8NGI9)	OR5A2	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5B2_HUMAN	(Q96R09)	OR5B2	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5B3_HUMAN	(Q8NH48)	OR5B3	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5BC_HUMAN	(Q96R08)	OR5B12	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5BH_HUMAN	(Q8NGF7)	OR5B17	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5DG_HUMAN	(Q8NGK9)	OR5D16	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5M8_HUMAN	(Q8NGP6)	OR5M8	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5W2_HUMAN	(Q8NH69)	OR5W2	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR8K1_HUMAN	(Q8NGG5)	OR8K1	11	11q12.1	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR5J2_HUMAN	(Q8NH18)	OR5J2	11	11q12.3	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
OR6A2_HUMAN	(O95222)	OR6A2	11	11p15	Olfactory II fam 6/MOR103 - 105, 107 - 119	S20
OR6Q1_HUMAN	(Q8NGQ2)	OR6Q1	11	11q12.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6M1_HUMAN	(Q8NGM8)	OR6M1	11	11q24.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6T1_HUMAN	(Q8NGN1)	OR6T1	11	11q24.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6X1_HUMAN	(Q8NH79)	OR6X1	11	11q24.1	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR5R1_HUMAN	(Q8NH85)	OR5R1	11	11q11	Olfactory II fam 8/MOR161 - 171	O13, O17, S20
OR8H1_HUMAN	(Q8NGG4)	OR8H1	11	11q11	Olfactory II fam 8/MOR161 - 171	O13, O17
OR8H2_HUMAN	(Q8N162)	OR8H2	11	11q11	Olfactory II fam 8/MOR161 - 171	O13, O17
OR8H3_HUMAN	(Q8N146)	OR8H3	11	11q11	Olfactory II fam 8/MOR161 - 171	O13
OR8K3_HUMAN	(Q8NH51)	OR8K3	11	11q11	Olfactory II fam 8/MOR161 - 171	O13
OR8K5_HUMAN	(Q8NH50)	OR8K5	11	11q11	Olfactory II fam 8/MOR161 - 171	O13

Continued

OR8U1_HUMAN	(Q8NH10)	OR8U1	11	11q11	Olfactory II fam 8/MOR161 - 171	O13
OR8I2_HUMAN	(Q8N0Y5)	OR8I2	11	11q12.1	Olfactory II fam 8/MOR161 - 171	O13
OR8J1_HUMAN	(Q8NGP2)	OR8J1	11	11q12.1	Olfactory II fam 8/MOR161 - 171	O13
OR8J3_HUMAN	(Q8NGG0)	OR8J3	11	11q12.1	Olfactory II fam 8/MOR161 - 171	O13
OR8B2_HUMAN	(Q96RD0)	OR8B2	11	11q24.1	Olfactory II fam 8/MOR161 - 171	O13
OR8B3_HUMAN	(Q8NGG8)	OR8B3	11	11q24.1	Olfactory II fam 8/MOR161 - 171	O13
OR8B4_HUMAN	(Q96RC9)	OR8B4	11	11q24.1	Olfactory II fam 8/MOR161 - 171	O13
OR8D1_HUMAN	(Q8WZ84)	OR8D1	11	11q24.1	Olfactory II fam 8/MOR161 - 171	O13
OR8D2_HUMAN	(Q9GZM6)	OR8D2	11	11q24.1	Olfactory II fam 8/MOR161 - 171	O13
OR8D4_HUMAN	(Q8NGM9)	OR8D4	11	11q24.1	Olfactory II fam 8/MOR161 - 171	O13
OR8A1_HUMAN	(Q8NGG7)	OR8A1	11	11q24.2	Olfactory II fam 8/MOR161 - 171	O13
OR8B8_HUMAN	(Q15620)	OR8B8	11	11q24.2	Olfactory II fam 8/MOR161 - 171	O13
OR8BC_HUMAN	(Q8NGG6)	OR8B12	11	11q24.2	Olfactory II fam 8/MOR161 - 171	O13
OR8G1_HUMAN	(Q15617)	OR8G1	11	11q24.2	Olfactory II fam 8/MOR161 - 171	O13
OR9G1_HUMAN	(Q8NH87)	OR9G1	11	11q11	Olfactory II fam 9/MOR120	S20
P2RY6_HUMAN	(Q15077)	P2RY6	11	11q13.5	Purinoceptor P2RY1 - 4, 6, 11 GPR91	A4, D18, S20
P2RY2_HUMAN	(P41231)	P2RY2	11	11q13.5 - q14.1	Purinoceptor P2RY1 - 4, 6, 11 GPR91	A4
GP152_HUMAN	(Q8TDT2)	GPR152	11	11q13.2	Putative/unclassified Class A GPCRs	S20
ADMR_HUMAN	(O15218)	ADMR	12	12q13.3	Adrenomedullin (G10D)	A4, A14
C3AR_HUMAN	(Q16581)	C3AR1	12	12p13.31	C5a anaphylatoxin	A3, A14
FZD10_HUMAN	(Q9ULW2)	FZD10	12	12q24.33	frizzled Group A (Fz 1 & 2 & 4 & 5 & 7 - 9)	F22
LGR5_HUMAN	(O75473)	LGR5	12	12q22 - q23	LGR like (hormone receptors)	A1
O10AD_HUMAN	(Q8NGE0)	OR10AD1	12	12q13.11	Olfactory II fam 10/MOR263 - 269	O7
O10A7_HUMAN	(Q8NGE5)	OR10A7	12	12q13.13	Olfactory II fam 10/MOR263 - 269	O7
O10P1_HUMAN	(Q8NGE3)	OR10P1	12	12q13.13	Olfactory II fam 10/MOR263 - 269	O7
OR6C2_HUMAN	(Q9NZP2)	OR6C2	12	12q13.13	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6C4_HUMAN	(Q8NGE1)	OR6C4	12	12q13.13	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6C1_HUMAN	(Q96RD1)	OR6C1	12	12q13.2	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR6C3_HUMAN	(Q9NZP0)	OR6C3	12	12q13.2	Olfactory II fam 6/MOR103 - 105, 107 - 119	O12
OR8S1_HUMAN	(Q8NH09)	OR8S1	12	12q13.11	Olfactory unclassified class II	O12, O17
RAI3_HUMAN	(Q8NFJ5)	GPRC5A	12	12p13 - p12.3	Orphan GPRC5	C16
GPC5D_HUMAN	(Q9NZD1)	GPRC5D	12	12p13.3	Orphan GPRC5	C16, D18
T2R43_HUMAN	(P59537)	TAS2R43	12	12p13.33	Taste receptors T2R	T23
V1AR_HUMAN	(P37288)	AVPR1A	12	12q14 - q15	Vasopressin type 1	A1
CLTR2_HUMAN	(Q9NS75)	CYSLTR2	13	13q14.12 - q21.1	Cysteinyl leukotriene	A4, A14
EDNRB_HUMAN	(P24530)	EDNRB	13	13q22	Endothelin	A1
LGR8_HUMAN	(Q8WXD0)	LGR8	13	13q13.1	LGR like (hormone receptors)	T23

Continued

5HT2A_HUMAN	(P28223)	HTR2A	13	13q14 - q21	Serotonin type 2	A2
BKRB1_HUMAN	(P46663)	BDKRB1	14	14q32.1 - q32.2	Bradykinin	A3, A14
BKRB2_HUMAN	(P30411)	BDKRB2	14	14q32.1 - q32.2	Bradykinin	A3, D18
LT4R1_HUMAN	(Q15722)	LTB4R	14	14q11.2 - q12	Leukotriene B4 receptor BLT1	A3, A14
LT4R2_HUMAN	(Q9NPC1)	LTB4R2	14	14q11.2 - q12	Leukotriene B4 receptor BLT2	A3
O10G2_HUMAN	(Q8NGC3)	OR10G2	14	14q11.2	Olfactory II fam 10/MOR263 - 269	O7
O10G3_HUMAN	(Q8NGC4)	OR10G3	14	14q11.2	Olfactory II fam 10/MOR263 - 269	O7
OR4K1_HUMAN	(Q8NGD4)	OR4K1	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4K2_HUMAN	(Q8NGD2)	OR4K2	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4K5_HUMAN	(Q8NGD3)	OR4K5	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4KD_HUMAN	(Q8NH42)	OR4K13	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4KE_HUMAN	(Q8NGD5)	OR4K14	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4L1_HUMAN	(Q8NH43)	OR4L1	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4M1_HUMAN	(Q8NGD0)	OR4M1	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4N2_HUMAN	(Q8NGD1)	OR4N2	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4N5_HUMAN	(Q8IXE1)	OR4N5	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4Q3_HUMAN	(Q8NH05)	OR4Q3	14	14q11.2	Olfactory II fam 4/MOR225 - 248	O10
O5AU1_HUMAN	(Q8NGC0)	OR5AU1	14	14q11.2	Olfactory II fam 5/MOR172 - 224, 249, 254	O11
PE2R2_HUMAN	(P43116)	PTGER2	14	14q22	Prostaglandin E2/D2 subtype EP2	A1
PD2R_HUMAN	(Q13258)	PTGDR	14	14q22.1	Prostaglandin E2/D2 subtype EP2	A1
SSR1_HUMAN	(P30872)	SSTR1	14	14q13	Somatostatin type 1	A3, A14
ACM5_HUMAN	(P08912)	CHRM5	15	15q26	Musc. acetylcholine Vertebrate type 5	T23, T24
OR4M2_HUMAN	(Q8NGB6)	OR4M2	15	15q11.2	Olfactory II fam 4/MOR225 - 248	O10
OR4N4_HUMAN	(Q8N0Y3)	OR4N4	15	15q11.2	Olfactory II fam 4/MOR225 - 248	O10
O4F15_HUMAN	(Q8NGB8)	OR4F15	15	15q26.3	Olfactory II fam 4/MOR225 - 248	O10
OR4F4_HUMAN	(Q96R69)	OR4F4	15	15q26.3	Olfactory II fam 4/MOR225 - 248	O10
OR4F6_HUMAN	(Q8NGB9)	OR4F6	15	15q26.3	Olfactory II fam 4/MOR225 - 248	O10
OR1F1_HUMAN	(O43749)	OR1F1	16	16p13.3	Olfactory II fam 1/MOR125 - 138, 156	O6
OR2C1_HUMAN	(O95371)	OR2C1	16	16p13.3	Olfactory II fam 2/MOR256 - 262, 270 - 285	T24
GPC5B_HUMAN	(Q9NZH0)	GPRC5B	16	16p12	Orphan GPRC5	T23
SSR5_HUMAN	(P35346)	SSTR5	16	16p13.3	Somatostatin type 5	T23
AA2BR_HUMAN	(P29275)	ADORA2B	17	17p12 - p11.2	Adenosine type 2	A2
CCR10_HUMAN	(P46092)	CCR10	17	17q21.1 - q21.3	C-C Chemokine type 10	A3
CCR7_HUMAN	(P32248)	CCR7	17	17q12 - q21.2	C-C Chemokine type 7	A3
FZD2_HUMAN	(Q14332)	FZD2	17	17q21.1	frizzled Group A (Fz 1 & 2 & 4 & 5 & 7 - 9)	F22
GALR2_HUMAN	(O43603)	GALR2	17	17q25.3	Galanin	A3, A14
GLP2R_HUMAN	(O95838)	GLP2R	17	17p13.3	Glucagon	B15, D18

Continued

GLR_HUMAN	(P47871)	GCGR	17	17q25	Glucagon	T23
OR1D2_HUMAN	(P34982)	OR1D2	17	17p13 - p12|17p13.3	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1A1_HUMAN	(Q9P1Q5)	OR1A1	17	17p13.3	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1A2_HUMAN	(Q9Y585)	OR1A2	17	17p13.3	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1D5_HUMAN	(P58170)	OR1D5	17	17p13.3	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1E1_HUMAN	(P30953)	OR1E1	17	17p13.3	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1E2_HUMAN	(P47887)	OR1E2	17	17p13.3	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1G1_HUMAN	(P47890)	OR1G1	17	17p13.3	Olfactory II fam 1/MOR125 - 138, 156	O6
OR3A1_HUMAN	(P47881)	OR3A1	17	17p13.3	Olfactory II fam 3/MOR255	O17
OR4D1_HUMAN	(Q15615)	OR4D1	17	17q23.2	Olfactory II fam 4/MOR225 - 248	O10
Q4VBP0_HUMAN	(Q4VBP0)	SSTR2	17	17q24	Somatostatin type 2	A3
GALR1_HUMAN	(P47211)	GALR1	18	18q23	Galanin	A3, D18
HRH4_HUMAN	(Q9H3N8)	HRH4	18	18q11.2	Histamine type 4	T23
MC5R_HUMAN	(P33032)	MC5R	18	18p11.2	Melanocortin hormone	A1, A14
MC4R_HUMAN	(P32245)	MC4R	18	18q22	Melanocortin hormone	A1, D18
C5AR_HUMAN	(P21730)	C5AR1	19	19q13.3 - q13.4	C5a anaphylatoxin	A3, D18
C5ARL_HUMAN	(Q9P296)	GPR77	19	19q13.33	C5a anaphylatoxin	T23, T24
CD97_HUMAN	(P48960)	CD97	19	19p13	EMR1	T23, T24
EMR2_HUMAN	(Q9UHX3)	EMR2	19	19p13.1	EMR1	B15
EMR3_HUMAN	(Q9BY15)	EMR3	19	19p13.1	EMR1	B15, D18
FPRL1_HUMAN	(P25090)	FPRL1	19	19q13.3 - q13.4	Fmet-leu-phe	A3, A14
FPRL2_HUMAN	(P25089)	FPRL2	19	19q13.3 - q13.4	Fmet-leu-phe	A3
FPR1_HUMAN	(P21462)	FPR1	19	19q13.4	Fmet-leu-phe	T23
FFAR1_HUMAN	(O14842)	FFAR1	19	19q13.1	Free fatty acid receptor (GP40, GP41, GP43)	D18
OR1I1_HUMAN	(O60431)	OR1I1	19	19p13.12	Olfactory II fam 1/MOR125 - 138, 156	O6
OR1M1_HUMAN	(Q8NGA1)	OR1M1	19	19p13.2	Olfactory II fam 1/MOR125 - 138, 156	O6
O10H1_HUMAN	(Q9Y4A9)	OR10H1	19	19p13.1	Olfactory II fam 10/MOR263 - 269	O7
O10H2_HUMAN	(O60403)	OR10H2	19	19p13.1	Olfactory II fam 10/MOR263 - 269	O7
O10H3_HUMAN	(O60404)	OR10H3	19	19p13.1	Olfactory II fam 10/MOR263 - 269	O7
O10H4_HUMAN	(Q8NGA5)	OR10H4	19	19p13.12	Olfactory II fam 10/MOR263 - 269	O7
O10H5_HUMAN	(Q8NGA6)	OR10H5	19	19p13.12	Olfactory II fam 10/MOR263 - 269	O7
OR2Z1_HUMAN	(Q8NG97)	OR2Z1	19	19p13.2	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
O4F17_HUMAN	(Q8NGA8)	OR4F17	19	19p13.3	Olfactory II fam 4/MOR225 - 248	O10
OR7A5_HUMAN	(Q15622)	OR7A5	19	19p13.1	Olfactory II fam 7/MOR139 - 155	O17
PE2R1_HUMAN	(P34995)	PTGER1	19	19p13.1	Prostaglandin E2 subtype EP1	A1, A14
PAR4_HUMAN	(Q96RI0)	F2RL3	19	19p12	Proteinase-activated	T23, T24
P2Y11_HUMAN	(Q96G91)	P2RY11	19	19p13.2	Purinoceptor P2RY1-4, 6, 11 GPR91	A4

Continued

EDG5_HUMAN	(O95136)	EDG5	19	19p13.2	Sphingosine 1-phosphate Edg-5	A1
EDG6_HUMAN	(O95977)	EDG6	19	19p13.3	Sphingosine 1-phosphate Edg-6	A1
EDG8_HUMAN	(Q9H228)	EDG8	19	19p13.2	Sphingosine 1-phosphate Edg-8	T23, T24
VN1R1_HUMAN	(Q9GZP7)	VN1R1	19	19q13.4	Vomeronasal receptors V1RL	D18
VN1R2_HUMAN	(Q8NFZ6)	VN1R2	19	19q13.42	Vomeronasal receptors V1RL	T23, T24
ADA1D_HUMAN	(P25100)	ADRA1D	20	20p13	Alpha Adrenoceptors type 1	A2
NPBW2_HUMAN	(P48146)	NPBWR2	20	20q13.3	GPR	T23, T24
HRH3_HUMAN	(Q9Y5N1)	HRH3	20	20q13.33	Histamine type 3	A2
NTR1_HUMAN	(P30989)	NTSR1	20	20q13 - 20q13	Neurotensin	A2, D18
OPRX_HUMAN	(P41146)	OPRL1	20	20q13.33	Opioid type X	T23, T24
SSR4_HUMAN	(P31391)	SSTR4	20	20p11.2	Somatostatin type 4	A3
AA2AR_HUMAN	(P29274)	ADORA2A	22	22q11.23	Adenosine type 2	A2
GALR3_HUMAN	(O60755)	GALR3	22	22q13.1	Galanin	T23
Q53ZR7_HUMAN	(Q53ZR7)	SSTR3	22	22q13.1	Somatostatin type 3	A3
CELR1_HUMAN	(Q9NYQ6)	CELSR1	22	22q13.3	Cadherin EGF LAG (CELSR)	B15, D18
OR2G2_HUMAN	(Q8NGZ5)	OR2G2	1	-	Olfactory II fam 2/MOR256 - 262, 270 - 285	O9
OR4E2_HUMAN	(Q8NGC2)	OR4E2	14	-	Olfactory II fam 4/MOR225 - 248	O10
AGTR2_HUMAN	(P50052)	AGTR2	X	Xq22 - q23	Angiotensin type 2	A3, A14
GRPR_HUMAN	(P30550)	GRPR	X	Xp22.2 - p22.13	Bombesin	A1, D18
BRS3_HUMAN	(P32247)	BRS3	X	Xq26 - q28	Bombesin	T23, T24
CXCR3_HUMAN	(P49682)	CXCR3	X	Xq13	C-X-C Chemokine type 3	A3
CLTR1_HUMAN	(Q9Y271)	CYSLTR1	X	Xq13.2 - 21.1	Cysteinyl leukotriene	A4
O13H1_HUMAN	(Q8NG92)	OR13H1	X	Xq26.2	Olfactory II fam 13/MOR253	O8
P2RY4_HUMAN	(P51582)	P2RY4	X	Xq13	Purinoceptor P2RY1 - 4, 6, 11 GPR91	T23
P2RY8_HUMAN	(Q86VZ1)	P2RY8	X|Y	Xp22.33; Yp11.3	Purinoceptor P2RY5, 8, 9, 10 GPR35, 92, 174	D18
P2Y10_HUMAN	(O00398)	P2RY10	X	Xq21.1	Purinoceptor P2RY5, 8, 9, 10 GPR35, 92, 174	T23
OPSG_HUMAN	(P04001)	OPN1MW	X	Xq28	Rhodopsin Vertebrate type 2	A1
OPSR_HUMAN	(P04000)	OPN1LW	X	Xq28	Rhodopsin Vertebrate type 2	A1
5HT2C_HUMAN	(P28335)	HTR2C	X	Xq24	Serotonin type 2	A2
GP173_HUMAN	(Q9NS66)	GPR173	X	Xp11	SREB	T23, T24
Other sapiens						
EDNRB_BOVIN	P28088	NM_174309.2				D21
OPSD_BOVIN	P02699	NM_001014890.1				D21
NK2R_BOVIN	P05363	NM_174469.2				D21
S1PR2_DANRE	Q9I8K8	XM_677820.3				D21
GP173_DANRE	Q9I918	NM_131498.1				D21
OR1A_DROME	Q9W5G6	NM_080290.3				D21

Continued

SMO_DROME	P91682	NM_078719.2	D21
MTH8_DROME	Q9W0V7	NM_138154.2	D21
ACM2_CHICK	P30372	NM_001030765.1	D21
GP175_CHICK	Q5ZII3	NM_001030599.1	D21
TRFR_CHICK	O93603	NM_204930.1	D21
SCTR_HUMAN	P47872	NM_002980.2	D21
NPY2R_HUMAN	P49146	NM_000910.2	D21
OR1_ANOGA	Q8WTE7	XM_318674.1	D21
OPSD_DICLA	Q9YGZ4	Y18673.1	D21
ANR_EISFO	Q75W84	AB121771.1	D21
FSHR_EQUAS	Q95179	U73659.1	D21
MSHR_EULFU (4, 1)	Q864F6	AY205141.1	D21
GPR56_PANTR	Q50DM7	NM_001114166.1	D21
BACR_HALSA	P02945	NC_002607.1	D21
AGTR1_BOVIN	P25104	NM_174233.2	T24
Q3T181_BOVIN	Q3T181	NM_001034738.1	T24
FSHR_BOVIN	P35376	NM_174061.1	T24
GNRHR_CANFA	Q9MZI6	NM_001003121.1	T24
EDNRA_CANFA	Q5KSU9	NM_001031632.1	T24
5HT1A_CANFA	Q6XXX9	NM_001012397.1	T24
OPS5_DROME	P91657	NM_057748.4	T24
GR33A_DROME	Q9VKA5	NM_080363.5	T24
MTH4_DROME	Q9V817	NM_166227.2	T24
OLF6_CHICK	P37072	NM_001031544.1	T24
FZD7_CHICK	O57329	NM_204221.1	T24
P2RY1_CHICK	P34996	NM_205333.1	T24
TAAR6_MOUSE	Q5QD13	NM_001010828.1	T24
NMBR_MOUSE	O54799	NM_008703.2	T24
FFAR2_MOUSE	Q8VCK6	NM_146187.3	T24
MLO10_ARATH	Q9FKY5	NM_125994.1	T24
5HT2A_MACMU	P50128	NM_001032966.1	T24
OPSD_MESBI	O62793	AF055316	T24
V1AR_MICOH	Q9WTV9	AF069304.2	T24