False positive non-synonymous polymorphisms of G-protein coupled receptor genes

Kersten M. Small, Carrie A. Seman, Alex Castator, Kari M. Brown, Stephen B. Liggett*

Departments of Medicine and Molecular Genetics, University of Cincinnati College of Medicine, 231 Albert Sabin Way, Room G062, Cincinnati, OH 45267-0564, USA

Received 15 February 2002; accepted 27 February 2002

First published online 13 March 2002

Edited by Julio Celis

Abstract Polymorphisms of G-protein coupled receptor (GPCR) genes are associated with disease risk and modification, and the response to receptor-directed therapy. Genomic sequencing (~1700 automated runs) from as many as 120 chromosomes from 60 multiethnic individuals was performed to confirm non-synonymous coding polymorphisms reported in the dbSNP database from 25 randomly selected GPCR genes. These polymorphisms were in regions of the receptors responsible for structural integrity, ligand binding, G-protein coupling and phosphoregulation. However, most of these putative polymorphisms could not be confirmed (false positive rate of 68%). Based on these results, we suggest that the variability of the superfamily is not well defined, and we caution against exclusive reliance on databases for selection of candidate GPCR polymorphisms for disease association and pharmacogenetic studies. © **2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.**

Key words: SNP; Database; Superfamily; Signaling

1. Introduction

G-protein coupled receptors (GPCRs) represent one of the largest families of genes in the genome, and carry out a diverse array of signaling including autocrine, hormonal, paracrine and central and peripheral neurotransmission [1]. As a superfamily, GPCRs provide integrated, dynamic, and flexible signaling to maintain homeostasis and to respond to physiologic and pathologic events. As such, pharmacologic agents which serve to activate, or antagonize, GPCR signaling are common. Indeed, ~60% of therapeutics currently target GPCRs [2].

The signaling of GPCRs in vivo and in vitro (from extracted tissues), as well as the clinical response to agonists and antagonists, displays a large degree of interindividual variation [3–6]. For example, treatment of congestive heart failure with β-adrenergic receptor antagonists has been found to be highly variable, with no consistent clinical parameter being predictive of a favorable response [4]. Some of the variability observed with GPCR agonists or antagonists is thought to be due to polymorphisms of the genes encoding these receptors. In asthma, the response to the β-agonist albuterol acting at airway β_2-adrenergic receptors is also highly variable [7], and calculations have revealed that ~50% of this variability has a genetic basis [6]. Besides potential pharmacogenetic loci, GPCR polymorphisms have also been found to be low-level risk factors for disease [8], and to act as disease modifiers [9]. While initial attention has focused on non-synonymous polymorphisms in the coding region of genes (i.e. polymorphisms that alter the encoded amino acid), it is clear that polymorphisms in promoter and untranslated regions may be responsible for altered expression or regulation as well [10].

Given such variability, the role of GPCRs in normal and pathologic states, and their use as targets for therapeutics, an understanding of the degree of genetic variability within their genes is necessary. In the process of carrying out recent studies on the signaling properties of polymorphic GPCRs, we noted that the available public databases which catalog polymorphisms appeared to have a high false positive rate. In the current study, we have utilized an adequately powered, multiethnic index repository of genomic DNA, and a set of randomly selected GPCRs to ascertain the validity of polymorphisms reported in the database dbSNP, the largest such collection of single nucleotide polymorphisms. The focus was on non-synonymous coding block variants, since structure/function studies have defined functional regions of many GPCRs, and thus implications of the genetic variability could be inferred.

2. Materials and methods

2.1. GPCR genes interrogated

On 8 August 2001 we delineated 142 GPCRs from the genome with known ligands (i.e. not orphan receptors). The opsins and the olfactory receptors were also excluded. Of these, 119 had a human sequence in the National Center for Biotechnology Information (NCBI) database with an entry in the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP) [11]. These genes were ordered using a random number generator and the coding regions of each examined for sequence variation in dbSNP. In order for the study to adequately represent the superfamily, coding region genes containing either one or two non-synonymous polymorphisms (but not more than two) were selected for study until 25 GPCR genes (~20% of the 119 characterized human GPCRs) were identified. This resulted in 36 polymorphisms reported in dbSNP which we targeted for verification. The non-synonymous nature of the polymorphism was confirmed by genomic reconstruction and translation using the indicated NCBI reference sequence (RefSeq).

2.2. Genomic DNA samples

Genomic DNA derived from 60 individuals was used for the study and consisted of samples from 20 Caucasians, 20 African-Americans, and 20 Asians. The samples were obtained from the Coriell Institute for Medical Research (Camden, NJ, USA). Using the equation:

*Corresponding author. Fax: (1)-513-558 0835.
E-mail address: stephen.liggett@uc.edu (S.B. Liggett).

Abbreviations: GPCR, G-protein coupled receptor

0014-5793/02/$22.00
PII: S0014-5793(02)02564-4

$P = 1-((1-f)^{2n})$

where P is the probability of detecting a polymorphism, f is the frequency of the variant, and n is the number of individuals sampled ($2n$ = number of chromosomes), it can be calculated that DNA samples from 60 individuals provides a 95% probability of detecting polymorphisms with allele frequencies as low as 0.025, and a 90% probability of detection at allele frequencies as low as 0.018. The samples were initially sequenced within a 30-sample cohort of 10 Caucasians, 10 African-Americans, and 10 Asians. If the polymorphisms provided by dbSNP were found, then no further sequencing was carried out. If not, the additional 30 samples were sequenced. Unequivocal sequence could not be obtained for two of the 36 SNPs. Thus, the results discussed represent 34 reported polymorphisms from 25 GPCR genes.

2.3. Sequencing

PCR primers were designed based on the reference sequence using the program Amplify 1.2 (Madison, WI, USA) with the expected polymorphism to be within 100–200 bases of the amplified sequence start site. PCR products were 300–400 bp in size. Primers also included sequences corresponding to M13 forward (−21) and M13 reverse universal sequencing primers so that the same sequencing primers could be used for all genes. Such an approach also provides for dye-primer sequencing chemistry, which in our experience results in higher quality data than dye-terminator sequencing. PCR reactions were optimized using the PCR optimizer kit (Invitrogen Life, Carlsbad, CA, USA) and included ~100 ng genomic DNA, 5 pmol of each sense and antisense primers, 0.8 nM dNTPs, 5 μl 5× buffer, ±10% DMSO, and 0.5 U platinum taq DNA polymerase in 20 μl reaction volume. PCR cycling started at 94°C for 4 min, followed by 35 cycles at 94°C for 30 s, 55–65°C for 30 s, and 72°C for 30 s, with a final extension at 72°C for 7 min. PCR products were diluted to 5 ng/μl prior to dye-primer cycle sequencing (ABI Prism, Foster City, CA, USA). For each PCR product, four separate reactions were performed to obtain sequence. Reactions consisted of 0.4 or 0.8 pmol of each dye-labeled primer (FAM to use with d/ddCTP mix, JOE to use with d/ddATP mix, TAMRA to use with d/ddGTP mix, and ROX to use with d/ddTTP mix), 1 μl of the appropriate d/ddNTP mix, 5 ng DNA, 0.625 U Amplitaq DNA polymerase FS, in a buffer containing 80 mM Tris–HCl, pH 9.0 and 2 mM $MgCl_2$. Reactions were subjected to 15 cycles at 96°C for 10 min, 55°C for 5 s, and 70°C for 60 s, followed by ethanol precipitation. Samples were subsequently resuspended in deionized formamide and heated to 90°C for 2 min prior to loading onto a 6% polyacrylamide gel for analysis using the ABI Prism 377 DNA sequencer.

2.4. Polymorphism detection

Sequence passed quality control if there were unequivocal base calls at the putative polymorphic site and the 5′ and 3′ flanking sequence. (Base calls of >99% for the entire PCR product, equivalent to a PHRED score of 20, were typical.) In addition, each chromatogram was inspected individually for confirmation. Particular attention was paid to the potential for heterozygosity, which with this technique is readily revealed by overlapping peaks with ~1/2 the expected peak height and is in the context of a low background [12]. For initial orientation and polymorphism identification, alignment with the reference sequence was carried out using MacVector 7.0 (Oxford Molecular, Madison, WI, USA). If the putative polymorphism was not detected in the first 30 samples using forward sequencing reactions, then the additional 30 samples were sequenced using the reverse M13 primer.

Table 1
Non-synonymous G-protein coupled receptor polymorphisms interrogated from the NCBI dbSNP database

Gene symbol	Rs#	NCBI RefSeq	Nucleotide position	Confirmed
ADORA2A	4990	NM_000675	1180	−
ADORA2A	4530	NM_000675	430	−
AGTR1	1801021	NM_004835	1106	−
AGTR1	1064533	NM_004835	1245	−
AGTR2	5191	NM_000686	904	+
AGTR2	1042860	NM_000686	965	−
BDKRB2	1046248	NM_000623	40	+
BRS3	5232	NM_001727	306	−
BRS3	5234	NM_001727	634	−
CALCR	1801197	NM_001742	1377	+
CCKBR	1805001	NM_00731	956	−
DRD1	5331	NM_000794	595	−
DRD2	1800496	NM_000795	961	−
DRD2	1110977	NM_000795	1084	−
DRD4	1800443	NM_000797	581	+
DRD5	1800762	NM_000798	989	−
F2R	5893	NM_001992	840	−
GALNR	5376	NM_001480	1773	+
GALNR	5377	NM_001480	1797	+
HTR1D	6299	NM_000864	837	+
HTR1E	6303	NM_000865	1351	−
HTR2C	6318	NM_000868	796	+
IL8RB	1805038	NM_001557	646	−
MTNR1A	1800884	NM_005958	502	−
MTNR1A	1800885	NM_005958	192	−
OXTR	810568	NM_000916	1439	−
OXTR	237901	NM_000916	809	−
PTGER1	1057362	NM_000955	285	−
PTGER3	5670	NM_000957	736	−
SSTR3	229568	M96738.1	1329	+
SSTR5	169068	NM_001053	1004	+
TRHR	5774	D16845	42	−
TSHR	1991517	NM_000369	2281	+
TSHR	1054708	NM_000369	1901	−

Entries are the gene symbol from Human Genome Organization (HUGO), the dbSNP unique reference identifier (Rs#), the reference sequence (RefSeq) as identified by NCBI, and the nucleotide position of each polymorphism within the NCBI RefSeq. A plus sign (+) indicates that, in the current study, the polymorphism was confirmed, and a minus sign (−) indicates a lack of confirmation.

3. Results and discussion

Clinical, cell-based, and transgenic mouse models of polymorphic variation in a few well-studied GPCRs have revealed that alteration in receptor expression or function can have important implications for human disease in at least three broad categories: risk factors for disease, disease modification, and alteration of the response to receptor-directed therapy.

In the current report, we assess the utility of the dbSNP database as a resource for identification of non-synonymous polymorphisms in GPCR genes. A relatively large number of genomic DNA samples (representing as many as 120 chromosomes) from a multiethnic cohort were interrogated at 34 putative polymorphic sites identified in dbSNP in 25 GPCR coding regions. Ultimately, >1700 sequencing runs were analyzed. The 25 random GPCR genes investigated are shown in Table 1. As anticipated, they represent a diverse range of subclasses and signal transduction pathways. The approximate locations of the amino acids affected by the putative 34 non-synonymous polymorphisms within a prototypic GPCR is depicted in Fig. 1. Putative substitutions included those in domains responsible for ligand binding (transmembrane domains), G-protein coupling (second and third intracellular loops and proximal cytoplasmic tail), phosphorylation by kinases such as protein kinase A, protein kinase C, and GPCR kinases (third intracellular loop and cytoplasmic tail), and overall receptor integrity (intracellular and extracellular loops).

Potentially, then, in considering development of new agonists or antagonists for a given receptor, the presence of many of these polymorphisms would be of significant concern. Indeed, even single amino acid substitutions within conformationally sensitive domains have been shown to dramatically alter receptor properties. For example, a polymorphism of the β_2-adrenergic receptor gene which results in a conservative Thr to Ile substitution in the fourth transmembrane domain of the receptor alters high- and low-affinity agonist binding, with a significant loss of receptor function [13]. Similarly, a Gly to Arg substitution in a G_s coupling domain due to a polymorphism in the β_1-adrenergic receptor gene results in a marked enhancement of receptor coupling to G_s and stimulation of adenylyl cyclase [14]. The same potential phenotypes evoked by these putative polymorphisms in these GPCR genes could also have important implications relative to genetic risk factors for disease or disease modifiers.

However, we found that the majority of the coding non-synonymous polymorphisms of GPCRs in dbSNP that were

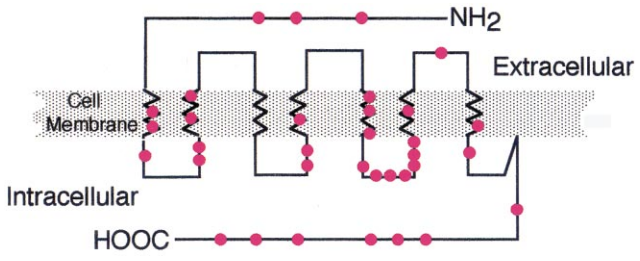

Fig. 1. Locations of the potential amino acid changes resulting from 34 putative SNPs within coding regions of GPCR genes as reported in the dbSNP database. Shown is a schematic of a prototypical GPCR, with the potential changed residues indicated as circles.

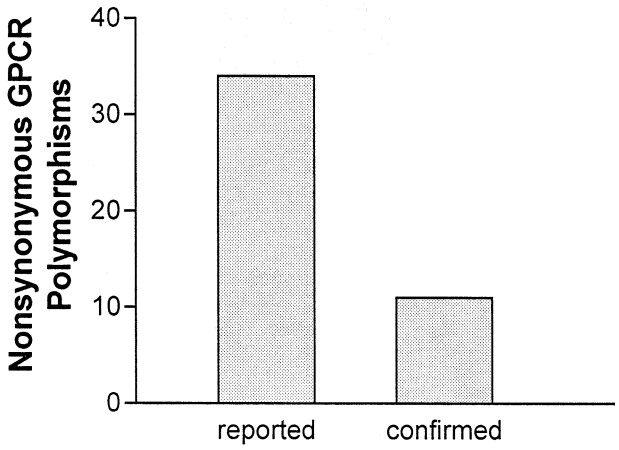

Fig. 2. Number of non-synonymous SNPs in 25 randomly selected GPCR genes reported in the dbSNP database compared to the number of polymorphisms confirmed in this study by sequence analysis. The false positive rate was 68%.

interrogated are likely erroneous. As shown in Fig. 2, only 32% of these putative polymorphisms could be confirmed (false positive rate of 68%). There was no relationship between the base context of the surrounding region and false positives (mean unconfirmed GC content of 56% vs. 52% GC content of confirmed polymorphisms). The number of chromosomes assayed by the submitter to dbSNP also did not appear to be a factor (mean 61 chromosomes for unconfirmed and 72 chromosomes for confirmed polymorphisms). There was a relationship between the allele frequency and whether a polymorphism was confirmed. In the confirmed group, when reported, the mean ($\pm$S.E.M.) allele frequency was 0.154 ± 0.054 vs. 0.043 ± 0.004 for the unconfirmed entries ($P = 0.02$).

The dbSNP database was established as a public resource in September 1998 by the NCBI, in collaboration with the National Human Genome Research Institute, as a general catalog of genome variation [11]. The contents of this database are cross-linked to records in other information resources such as GenBank, LocusLink, the human genome sequence, and Pub Med. Most ($>99\%$) of nucleotide sequence variations in this database are SNPs; however, small insertions/deletions, microsatellite repeat variation, and retrotransposable element insertions are represented [15]. At the time this study was initiated, dbSNP contained 1 805 951 unique human sequence variants.

Examples of GPCR polymorphisms that are risk factors for disease are the studies showing that the Arg16$\rightarrow$Gly and Gln27$\rightarrow$Gln polymorphisms of the β_2-adrenegic receptor are associated with hypertension [8]. These results are consistent with the altered vasoreactivity to β-agonists of individuals with certain β_2-adrenergic receptor genotypes [16,17], as well as the in vitro studies in transfected cells [18,19]. In one of the most comprehensive studies to date, Boerwinkle and colleagues [8] have shown that the odds ratio for the occurrence of hypertension for the Glu27 allele was 1.80. This relatively low level of risk may be typical of common polymorphisms of individual GPCR genes, particularly in complex diseases, which likely involve multiple genes as well as environmental factors.

GPCR polymorphisms have been associated with an altered clinical phenotype in several diseases. For instance, the β_1-

adrenergic receptor Ser49→Gly polymorphism imparts enhanced agonist-promoted downregulation in transfected cells [20]. In patients with congestive heart failure, the Gly49 receptor appears to provide a protective effect [21], with ~50% lower five-year mortality compared to those with Ser49. This is consistent with the notion that β_1-adrenergic downregulation is protective against the high levels of catecholamines that typically occur with the syndrome.

Common polymorphisms which alter the response to drugs have the potential to segregate populations so as to optimize therapy. Such pharmacogenetic loci are well defined for certain drug metabolism enzymes [22], but the field is only beginning to be explored with GPCR polymorphisms. The altered bronchodilatory response to β-agonist, and tachyphylaxis to its continued use, in the treatment of asthma appears to be influenced by β_2-adrenergic receptor coding and 5′ upstream polymorphisms [7,10].

In this study we have assessed the usefulness of the dbSNP database for identification of non-synonymous coding polymorphisms of GPCRs. Of note, the relatively low allele frequencies for these non-synonymous polymorphisms of GPCR genes are consistent with data from random genes [23] which show that ~60% of all non-synonymous polymorphisms have allele frequencies of <0.05. This prevalence of low-allele-frequency non-synonymous polymorphisms further emphasizes the need for highly accurate databases. Our findings from the current study indicate a high number of false positive polymorphisms in this group of important signaling genes in dbSNP, which amounts to a positive predictive value of only 32%. It thus appears that polymorphisms identified *only* by dbSNP may not be reliable candidates for resequencing in clinical studies. Such efforts might result in negative association studies and the given GPCR gene not further considered. This could represent a significant error in gene selection for such studies. In addition, while we have no way of assessing false negative GPCR polymorphisms in the database, they may be common as well, particularly for non-synonymous polymorphisms where the allele frequency is low. Of note, our study is confined to GPCR non-synonymous SNPs with relatively low allele frequencies, and may not be applicable to polymorphisms in other genes or regions of genes, particularly if allele frequencies are higher [24].

We recognize the tremendous accomplishments of the Human Genome Project [25], and the substantial contributions of efforts aimed at understanding human genetic variability that have been associated with this and other efforts [23,26,27]. Our current report thus represents a cautionary note that at least within the GPCR superfamily, there is a high degree of false positive non-synonymous polymorphisms in the dbSNP public database. As such, the extent or degree of genetic variation in the superfamily is not well defined. And, our results emphasize that when considering a GPCR gene for polymorphisms, diligence in resequencing the entire region(s) of interest in a multiethnic cohort of adequate size is necessary.

Acknowledgements: This work was funded by the National Institutes of Health Grant U01 GM61376 and ES06096. We wish to thank Esther Getz for assistance in manuscript preparation.

References

[1] Rana, B.K., Shiina, T. and Insel, P.A. (2001) Annu. Rev. Pharmacol. Toxicol. 41, 593–624.
[2] Muller, G. (2000) Curr. Med. Chem. 7, 861–888.
[3] Bristow, M.R., Hershberger, R.E., Port, J.D., Minobe, W. and Rasmussen, R. (1988) Mol. Pharmacol. 35, 295–303.
[4] van Campen, L.C., Visser, F.C. and Visser, C.A. (1998) J. Cardiovasc. Pharmacol. 32 (Suppl. 1), S-31–S-35.
[5] Liggett, S.B., Shah, S.D. and Cryer, P.E. (1989) Am. J. Physiol. 256, E-600–E-609.
[6] Drazen, J.M., Silverman, E.K. and Lee, T.H. (2000) Br. Med. Bull. 56, 1054–1070.
[7] Israel, E., Drazen, J.M., Liggett, S.B., Boushey, H.A., Cherniack, R.M., Chinchilli, V.M., Cooper, D.M., Fahy, J.V., Fish, J.E., Ford, J.G., Kraft, M., Kunselman, S., Lazarus, S.C., Lemanske, R.F., Martin, R.J., McLean, D.E., Peters, S.P., Silverman, E.K., Sorkness, C.A., Szefler, S.J., Weiss, S.T. and Yandava, C.N. (2000) Am. J. Respir. Crit. Care Med. 162, 75–80.
[8] Bray, M.S., Krushkal, J., Li, L., Ferrell, R., Kardia, S., Sing, C.F., Turner, S.T. and Boerwinkle, E. (2000) Circulation 101, 2877–2882.
[9] Turki, J., Pak, J., Green, S., Martin, R. and Liggett, S.B. (1995) J. Clin. Invest. 95, 1635–1641.
[10] Drysdale, C.M., McGraw, D.W., Stack, C.B., Stephens, J.C., Judson, R.S., Nandabalan, K., Arnold, K., Ruano, G. and Liggett, S.B. (2000) Proc. Natl. Acad. Sci. USA 97, 10483–10488.
[11] Sherry, S.T., Ward, M. and Sirotkin, K. (1999) Genome Res. 9, 677–679.
[12] Small, K.M., Rathz, D.A. and Liggett, S.B. (2002) Methods Enzymol. 343, 459–475.
[13] Green, S.A., Cole, G., Jacinto, M., Innis, M. and Liggett, S.B. (1993) J. Biol. Chem. 268, 23116–23121.
[14] Mason, D.A., Moore, J.D., Green, S.A. and Liggett, S.B. (1999) J. Biol. Chem. 274, 12670–12674.
[15] Sherry, S.T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) Nucleic Acids Res. 29, 308–311.
[16] Dishy, V., Sofowora, G.G., Xie, H.-G., Kim, R.B., Byrne, D.W., Stein, C.M. and Wood, A.J.J. (2001) N. Engl. J. Med. 345, 1030–1035.
[17] Hoit, B.D., Suresh, D.P., Craft, L., Walsh, R.A. and Liggett, S.B. (2000) Am. Heart J. 139, 537–542.
[18] Green, S., Turki, J., Innis, M. and Liggett, S.B. (1994) Biochemistry 33, 9414–9419.
[19] Liggett, S.B. (2002) N. Engl. J. Med. 346, 536–538.
[20] Rathz, D.A., Brown, K.M., Kramer, L.A. and Liggett, S.B. (2002) J. Cardiovasc. Pharmacol. 39, 155–160.
[21] Borjesson, M., Magnusson, Y., Hjalmarson, A. and Andersson, B. (2000) Eur. Heart J. 21, 1810–1812.
[22] Nebert, D.W. and Dieter, M.Z. (2000) Pharmacology 61, 124–135.
[23] Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q. and Lander, E.S. (1999) Nat. Genet. 22, 231–238.
[24] Marth, G., Yeh, R., Minton, M., Donaldson, R., Li, Q., Duan, S., Davenport, R., Miller, R.D. and Kwok, P.-Y. (2001) Nat. Genet. 27, 371–372.
[25] International Human Genome Sequencing Consortium (2001) Nature 409, 860–921.
[26] The International SNP Map Working Group (2001) Nature 409, 928–933.
[27] Stephens, J.C., Schneider, J.A., Tanguay, D.A., Choi, J., Acharya, T., Stanley, S.E., Jiang, R., Messer, C.J., Chew, A., Han, J.-H., Duan, J., Carr, J.L., Lee, M.S., Koshy, B., Kumar, A.M., Zhang, G., Newell, W.R., Windemuth, A., Xu, C., Kalbfleisch, T.S., Shaner, S.L., Arnold, K., Schulz, V., Drysdale, C.M., Nandabalan, K., Judson, R.S., Ruano, G. and Vovis, G.F. (2001) Science 293, 489–493.