

In Silico Prediction of Deleterious Non-Synonymous SNPs of Human GABRA2 gene and Altered Protein Structure and Function – A Link to Alcohol Dependence?

Sathish Sankar (✉ sathish3107@gmail.com)

Saveetha Dental College and Hospital

Subramanian Suganya

Auxilium College

Research Article

Keywords: SNP, GABRA2, GABAA, alcohol dependence, protein function

Posted Date: March 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1458373/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Alcohol Dependence is a serious and common public health problem around the globe. Genetic factors contribute to the risk of alcohol dependence. Mapping and identifying the specific genes that influence alcohol dependence and factors that alter the genetic expression are the thrust areas. Single-nucleotide polymorphisms (SNPs) in *GABRA2* are associated with impulsiveness-related traits. The present study looked at non-synonymous SNPs with the potential to affect the structural and functional properties of the protein. The missense variants in the *GABRA2* gene from sequences available in the NCBI database were analyzed using different bioinformatics tools. The modeled protein structures of the mutant proteins were compared with the native protein to determine stability changes. The identified deleterious variants were mostly present in the neurotransmitter-gated ion-channel ligand-binding domain and predicted to cause potential structural and functional changes in the protein. Among the 14 mutants, R58T was found to have the highest energy of -19646 KJ/mol, even after energy minimization, when compared with the native structure. The remaining 13 mutants were found to have lower energy values ranging from -13567 to -13857 KJ/mol and 10 mutants were found in the conserved regions of the protein sequence. With a better understanding of the genetic basis of alcohol dependence, it is possible to pre-screen 'at-risk' individuals and design personalized early intervention, especially among the youth population.

Introduction

Alcohol dependence is a complex disorder presenting protean clinical manifestations. This involves multifactorial influences like host genetics, and environmental influences (Nusbaumer and Reiling, 2002; Littlefield et al. 2010). *GABRA2* gene codes for the alpha-2 subunit of the GABA-A receptor, one of the ionotropic receptors which have been related to anxiety, depression and other behavioural disorders, including drug dependence and schizophrenia (Gonzalez-Nunez, 2015). Genetic variants or single-nucleotide polymorphisms (SNPs) in *GABRA2* have previously been shown to be associated with impulsiveness-related traits (Villafuerte et al. 2013). Impulsiveness is a behavioural risk factor for alcohol and other substance abuse. Host genetics plays a role in different impulsivity-related traits (Feldstein et al. 2009; Birkley and Smith, 2011). Genetic analysis screening for SNPs genome-wide has shown it is possible to identify genes involved in alcohol dependence through different biological pathways (Bierut et al. 2010). In particular, the non-synonymous SNPs (nsSNPs) or missense variants play a pivotal role as they are associated with changes in the translated protein sequence. This leads to the functional diversity of the encoded proteins and in particular associated with alcohol dependence (Way et al. 2017).

Protein function is altered by reduction of protein solubility or destabilization of protein structure due to nsSNPs. Such polymorphisms may affect gene regulation by altering transcription and translation. The variations in the alleles of alcohol dependence genes are shown to be associated with higher impulsive behaviour (Taylor et al. 2016; Huang et al. 2017). Repetitive genome-wide scans show the linkage of alcohol dependence to a region on chromosome 4p, which contains a cluster of genes encoding GABA-A receptor subunits (Covault et al. 2004; Tretlein et al. 2009). Linkage disequilibrium analyses of 69 SNPs within a cluster of 4 GABA-A receptor genes including *GABRA2* was reported (Edenberg et al. 2004). The study found 31 SNPs in *GABRA2*, one SNP was seen in the flanking genes and showed significant association with alcoholism. The other SNP was associated with brain oscillations in the beta frequency. The region of the *GABRA2* gene with the strongest association to alcohol dependence extended from intron 3. An interesting observation was that 43 of the consecutive 3-SNP haplotypes in this region of *GABRA2* were significantly associated with alcohol abuse (Begleiter and Porjesz, 2006).

This underscores the probable contribution of polymorphic variation at the *GABRA2* locus associated with the risk for alcohol dependence. Previously, the association between 11 variants (SNPs) in *GABRA2* with NEOimpulsiveness (altered personality traits) and drinking-related problems were reported (Villafuerte et al. 2013). Ten of these SNPs were associated in a statistically significant manner with NEOimpulsiveness. Clarke et al. (2017) reported eight independent loci associated with alcohol consumption. The association between alcohol consumption and alcohol metabolizing genes (ADH1B/ADH1C/ADH5) and the Beta-klotho gene (KLB) have been documented. As postulated previously by Hassan et al. 2016, missense variants play an important role by affecting the translated protein and leading to disease.

dbSNP is a database of sequence variations, which includes single nucleotide substitutions that could be frequent or rare in a given population. The clinical impact of these SNPs are unknown or not completely studied. Several SNPs are associated with a tendency towards alcoholism (<https://www.snpedia.com/index.php/Alcoholism>). The present study reported in the manuscript was carried

out to look at non-synonymous SNPs with the potential to affect the structural and functional properties of the protein using bioinformatic tools. We investigated missense variants in the GABRA2 gene from sequences available in the NCBI (National Center for Biotechnology Information) dbSNP database. Little is known about the role of nsSNPs in GABRA2 gene about the functional and structural stability changes of the protein. In our study, information on SNPs was obtained from databases and analyzed using different bioinformatics tools for inference on the role of GABRA2 receptors. As the majority of high-risk mutations affect protein stability, we also examined the modeled protein structures of the mutant proteins and compared them with the native protein to determine stability changes.

Materials And Methods

In this study, an *in silico* analysis was carried out by acquiring datasets from NCBI database, protein databank and analyzing using the software as indicated below:

Datasets:

The SNPs (n=7453) of human *GABRA2* gene coding *GABRA2* protein (NCBI Accession: NP_000798) was retrieved from the NCBI dbSNP database. The SNPs are considered to be deleterious when they are linked to disease states. The SNPs belonging to different functional classes were obtained from the database and are shown in Figure 1. Among the SNPs, 93 were missense variants, other SNPs occurred in intronic region (n=6811), 3' untranslated region (UTR) (n=329), 5'UTR (n=134), coding synonymous (n=72), non-sense variants (n=3), stop-gains (n=3) and frameshift variants (n=8). Only the missense variants were selected for further analysis. Validation of 93 missense SNPs was carried out using Ensembl and UCSC browsers. A total of 86 SNPs were selected for further analysis including prediction of their protein structure, stability and function.

In silico methods:

SIFT (Sorting Intolerant From Tolerant) program available at <https://sift.bii.a-star.edu.sg/> was used to predict the deleterious or damaging nature of the 86 missense SNPs. This program is based on sequence homology, physical properties of amino acids and the degree of evolutionary conservation of the sequence among various species.

SIFT predictions are given as either "damaging" or "tolerated". The former indicates that the substitution is predicted to affect protein function and the latter indicates that the substitution is predicted to be functionally neutral. A SIFT score of zero indicates evolutionary conserved and intolerance towards substitutions, while scores close to one indicate tolerance towards substitution. Scores <0.05 are predicted by the algorithm to be intolerant or highly deleterious while scores >0.05 are regarded as highly tolerant of substitutions. Each of the programs listed below was used to independently analyze the 86 missense SNPs in the *GABRA2* gene:

Polyphen (Polymorphism and Phenotyping) server available at <http://genetics.bwh.harvard.edu/pph2/> was used to screen and predict the deleterious nsSNPs that are based on the observable structural changes induced by the nsSNPs. PANTHER (Protein Analysis through Evolutionary Relationships) server available at <http://pantherdb.org> was used to calculate the duration of a given amino acid that has been evolutionarily preserved among various species and predicts the effect of the specific amino acid change on the structural and functional aspects of the protein. The longer the amino acid is conserved during evolution, the greater the likelihood of having functional importance in protein structure and function.

PROVEAN (Protein Variation Effect Analyzer) server available at <http://provean.jcvi.org/index.php> was used to predict if single or multiple indels and substitutions in the amino acid sequence affect protein function. The program utilizes clustering of BLAST hits with 75% global sequence identity. The top 30 clusters of closely related sequences are used to generate the prediction by the program. Each supporting sequence is assigned a delta alignment score which is then averaged within and across clusters to generate the PROVEAN score. The score ≤ -2.5 indicates the protein variant predicted has a "deleterious" effect.

Mutation Assessor program that predicts the functional impact of amino acid substitutions in proteins. In this program, the functional impact is assessed based on the evolutionary conservation of the affected amino acid in protein homologs. Prediction of pathological (disease-associated) mutations is carried out using PMut <http://mmb.irbbarcelona.org/PMut>. The final output is

displayed as a pathogenicity index ranging from 0 to 1 (indexes > 0.5 single pathological mutations) and a confidence index ranging from 0 (low) to 9 (high).

I-Mutant 3.0 (<http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi>), a support vector machine (SVM) based tool was used to assess the 86 SNPs in the *GABRA2* gene. It classifies the prediction as (i) neutral mutation ($-0.5 \leq \text{DDG} \leq 0.5$ kcal/mol) (ii) large decrease (≤ -0.5 kcal/mol) and (iii) large increase (> 0.5 kcal/mol). The protein stability changes impacted by single point mutations is predicted using this program. This program was used both as a classifier for predicting the sign of the protein stability change upon mutation and as a regression estimator for predicting the related change in Gibbs-free energy ($\Delta\Delta G$). I-Mutant Disease (Predictor of human Deleterious Single Nucleotide Polymorphisms) built within I-Mutant suite was used.

MutPred (<http://mutpred.mutdb.org/>) based upon SIFT algorithm and a gain/loss was used for 14 predicted structural and functional properties. MutPred is a web application tool developed to classify an amino acid substitution (AAS) as disease-associated or neutral in humans. In addition, it also predicts the molecular cause of disease/deleterious AAS.

Modeling of the *GABRA2* protein structure:

Crystallized protein structures of *GABRA* genes were not available and therefore, computational modeling based on homology prediction was utilized to construct the reference protein structures. I-TASSER web server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) was employed for 3D protein modeling. I-TASSER generated 5 models, of which the model that had the highest confidence score (C-score) along with RMSD (Root Mean Square Deviation) score and TM (Template Modeling)-score was selected for further analysis.

Superimposition of wild type - mutant proteins and RMSD calculation

Each mutant model (14 models) was generated using the “mutation tool” in Swiss-PDBViewer. The mutation tool was used to replace the native amino acid with the “best” rotamer of the new amino acid. Energy minimization for the predicted models was performed with the GROMOS 43B1 field implementation of Deep View v4.1 tool (https://spdbv.vital-it.ch/energy_tut.html). This force field was built to evaluate the energy of a protein structure as well as repair distorted geometries through energy minimization.

Energy minimization for both the native and the mutated protein models was carried out using this program. The RMSD values of the atoms upon superimposing the native and the mutant protein structure was calculated using Swiss-PDBViewer by the “Calculate RMS” function. The extent of structural deviation between the native and the mutant protein structures associated with a functional effect on the protein was predicted by calculating the RMSD by superimposing the native and protein structures. The higher the RMSD value, the structural deviation is more likely to be associated with the altered function of the protein. The stability of the mutant protein structure was then analyzed by the I-Mutant server (<http://folding.biofold.org/i-mutant/i-mutant2.0.html>).

Validation of the native and the mutant model using Ramachandran Plot

The Ramachandran Plot was used to calculate the dihedral angles of the amino acid residues and to predict the energetically allowed residues based upon their phi and psi dihedral angles, thereby ascertaining the structural and functional properties of the protein structure. The energy minimized native and the mutant protein models were validated with the online tool RAMPAGE program (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>).

ConSurf tool (<http://consurf.tau.ac.il/2016/>) was used to estimate the evolutionary conservation of amino acid positions in the *GABRA2* protein sequence. This analysis is based on phylogenetic relations between homologous sequences. The degree of conservation of amino acid residues was estimated using default program settings. The highly conserved residues were identified and the residues (exposed/buried) in the protein structure located at the sites of high-risk nsSNPs were identified. The conserved regions were predicted utilizing colouring scheme and conservation scores (conservation scores: 1–4 variables, 5–6 intermediate, and 7–9 conserved).

GeneMANIA, an online database (<http://www.genemania.org/>) that predicts the function of the genes and gene sets using a very large set of functional association data was used to analyze the 86 SNPs. The web interface generates hypotheses about gene function, analyzing gene lists and prioritizing genes for functional assays. A list of functionally similar genes identified using available genomics and proteomics data were generated using a very large set of functional association data.

Results

Fourteen nsSNPs were common to those predicted by all the six programs (SIFT, PolyPhen, PANTHER, PROVEAN, Mutation Assessor and PMut); these 14 SNPs are listed in Table 1. The details of the SNPs as identified by individual programs are shown in supplementary tables 1-6. Among the 14 nsSNPs, I-Mutant analysis predicted 13 nsSNPs to be associated with decreased stability, except one, rs765251624 (R58T) that showed an increase in the stability (Table 1). The variants that were predicted to have decreased stability were also found to have increased RMSD values. The cross-validation of the results of I-Mutant Disease in I-Mutant suite 3.0 showed the three variants (V208F, P426S, L170S) were neutral and the remaining variants were predicted to be associated with disease. The RMSD values were in keeping with the I-Mutant results. In the analysis, one variant (I255T) was marked as unknown and RI and DDG value were not obtained from the program.

A total of 86 nsSNPs were identified from the database and analyzed using different *in silico* programs. Of these, 29 nsSNPs were predicted to be functionally deleterious (affecting protein structure) by the SIFT server showing a highly deleterious tolerance index score of 0.00. The other remaining variants that were predicted as “tolerated”. Functionally deleterious nsSNPs as predicted by SIFT program are shown in supplementary table 1.

The PolyPhen server predicted 42 nsSNPs of 86 SNPs to be functionally deleterious to the protein structure. Of which, 29 nsSNPs were predicted to be “probably damaging” with the score ranging from 0.96 to 1.00 and 13 nsSNPs were predicted to be “possibly damaging” with the score ranging from 0.454 to 0.933 (supplementary table 2). The PANTHER server predicted 76 nsSNPs to be damaging and the remaining nsSNPs were predicted to be benign (supplementary table 3). Among 86 nsSNPs, 22 nsSNPs that were predicted as deleterious were also predicted as deleterious by the other three programs *viz.* SIFT, PolyPhen, and PANTHER server.

The PROVEAN server predicted 25 nsSNPs to be functionally damaging out of the 86 nsSNPs submitted for analysis (supplementary table 4). Of these, 18 nsSNPs were also predicted by SIFT, PolyPhen, and PANTHER servers. Mutation assessor generated 25 nsSNPs predicted to be associated with a diseased phenotype. Of those, 15 nsSNPs were also predicted by SIFT, PolyPhen, PANTHER, and PROVEAN servers (supplementary table 5).

The functional impact of 86 deleterious nsSNPs in *GABRA2* protein was analyzed using PMut server. Of the 86 nsSNPs, 37 are classified as pathological, and the remaining were neutral. Among those, 14 nsSNPs were common to those predicted by the above five servers (SIFT, PolyPhen, PANTHER, PROVEAN and Mutation Assessor) (supplementary table 6). MutPred was used to determine the tolerance degree for each amino acid substitution based on physio-chemical properties. The variants (P445H, P280S, I255T, V208F, S186C, L170S, W122L, I121N, F93C, Y73C, R58T and T43A) were predicted to cause potential structural and functional changes in the protein. A188T was predicted to cause less significant functional change. The results of the Mutpred prediction server are shown in Table 2.

Studies on protein mutation and energy minimization of the native and mutated protein

The 14 nsSNPs predicted to be potentially deleterious by all the 6 programs were mapped into the *GABRA2* protein using the “mutation tool” in Swiss-PDBViewer to replace the native amino acid with a new one. Energy minimization of both the native and each of the mutant proteins was done with the help of Swiss-PDBViewer. The resulting energy values of the native and the mutant structures are given in Table 1. The total energy of the native protein structure was determined to be -19946.123 KJ/mol. Among the

14 mutants, R58T was found to have the highest energy of -19646 KJ/mol, even after energy minimization, when compared with the native structure. The remaining 13 mutants were found to have lower energy values ranging from -13567 to -13857 KJ/mol.

The total energy comparison showed that mutants have lower energy values than the *GABRA2* protein. The difference in the RMSD value for each of the three mutant proteins compared to native protein were 0.00Å. The RMSD value of the other mutants was higher than that for the native protein ranging from 0.23Å to 0.15Å, indicating structural changes. The higher the RMSD value, the more will be the deviation between the native and the mutant protein structures; this, in turn, alters the protein's stability and functional activity. The native *GABRA2* protein as predicted by I-TASSER program is shown in Figure 2 highlighting the identified deleterious SNPs and the position in the Neurotransmitter-gated ion-channel ligand-binding domain. The superimposed mutant protein structures are shown in Figure 3.

Validation of the native and the mutant model using Ramachandran Plot.

The energy minimized native and mutant protein structures in .pdb formats were submitted to RAMPAGE for validating the protein structure using the Ramachandran plot. The results are shown in Figures 4 and 5. The native protein model contains 339 residues (75.5%) in the favoured region, 67 residues (14.9%) in the allowed region and 15 residues (5.3%) in the outlier region. The mutant protein models also showed similar results (Table 1) indicating that there are no major structural changes in mutant protein models compared to the native protein model.

Conservation analysis of *GABRA2* protein

ConSurf analysis identified conserved residues in *GABRA2* protein and predicted residues to be exposed or buried in the *GABRA2* protein structure (Fig6). ConSurf exploits evolutionary variation in multiple sequence alignments to determine the degrees of conservation. The results show that among predicted 14 deleterious nsSNPs, ten nsSNPs (T43A,R58T,F93C,I121N,W122L,S186C,A188T,I255T,P280S,P426S) occurred in conserved sites.

The results of biological interaction network analysis by GeneMania are shown in Figure 7. The *GABRA2* gene is predicted to interact with other genes and the major function was shown to have neurotransmitter receptor activity and neuron-neuron synaptic transmission. Figure 7 shows the gene-gene interactions of *GABRA2*. The most important interactions are with GABRA1, GABRA3, GABRA6, GABRA4, GABRG2, GABRA6 and GABRB2.

Discussion

The development of alcohol dependence is a complex and dynamic process that needs to be investigated in terms of comorbidities with psychiatric disorders especially to develop new psychotherapeutic and pharmacotherapeutic options (Farren et al. 2012). Studies support the importance of genetic influences in substance abuse and dependence. Some specific genes of interest are associated with alcohol use disorders (Mayfield et al. 2008). A study reported from Germany has documented *GABRA2* gene sequence variation concerning alcohol abuse. Of the four haplotypes investigated, T-CA-C-A-T-T-C haplotype was significantly more often present in alcohol-dependent subjects compared to controls (Soyka et al. 2008).

Strong associations between SNPs in the gene encoding the alpha2 subunit of the *GABRA2* was found with alcohol dependence and affecting brain oscillations as seen in distinct electroencephalography patterns (Edenberg et al. 2004). The link between alcohol abuse and SNPs in *GABRA2* was found in subjects with illicit drug dependence (Agrawal et al. 2006). Genetic variations particularly, nsSNPs resulting in amino acid changes disrupt potential functional sites responsible for protein activity, structure, or stability (Schaefer et al. 2012). Investigating the role of nsSNPs in the structural and functional changes in *GABRA2* will help in understanding the genetic mechanism of alcohol dependence associated with impulsive disorders. In the study on 295 Americans, of whom 97% of patients were of Caucasian origin, primarily the intronic regions of *GABRA2* gene were analyzed to detect association with impulsiveness and lifetime alcohol problems. Our study, looked at missense variants in exons collected from the NCBI database representing the diverse population data as this would directly affect the *GABRA2* protein stability and function.

The enormous human genomic sequence information obtained from large-scale projects are helpful in several computational approaches to identify the protein mutants in terms of single amino acid changes that disrupt gene functions. Several prediction tools have been developed to identify amino acid variants. The programs such as SIFT, PolyPhen-2, Mutation Assessor, MAPP, PANTHER, LogR.E-value, Condel and several others predict the effect of missense variants on protein function. The SIFT program prediction through PSI-BLAST indicated 29 nsSNPs as damaging with scores ≤ 0.05 . The 14 commonly predicted nsSNPs had the least score of zero indicating the high predictive ability of the program.

PolyPhen-2 uses eight sequence-based and three structure-based predictive features by comparing the property of the wild-type and the corresponding mutant allele that together defines an amino acid replacement (Adzhubei et al. 2010). Our SNP data were analyzed in terms of pph2_prob (classifier probability of the variation being damaging), pph2_FPR [classifier model False Positive Rate (1 - specificity) at the above probability] and pph2_TPR [classifier model True Positive Rate (sensitivity) at the above probability]. Among the 14 nsSNPs identified by all the six programs, the 13 nsSNPs had high pph2_prob compared to other predicted mutations ranging from 0.99 to 1 and were predicted as probably damaging.

The PANTHER program estimates the likelihood of a particular nonsynonymous coding SNP to cause a functional impact on the protein (Tang et al. 2016). The position-specific evolutionary preservation (PSEP) tool employed in the PANTHER program uses a distinct metric based on evolutionary preservation wherein it calculates the length of time (in millions of years) a given amino acid has been preserved in the lineage leading to the protein of interest. The longer the preservation time, the greater the likelihood of functional impact. The 76 SNPs that included the 14 nsSNPs were identified as probably damaging.

Out of the 86 nsSNPs, the PROVEAN server program predicted 25 nsSNPs to be functionally damaging, of which, 18 nsSNPs were predicted by other programs such as SIFT, PolyPhen, and PANTHER servers. The 14 nsSNPs were predicted by all six programs with scores ranging from -3.13 to -11.96. The prediction accuracy of this program for human protein variations was reported to be 79.5%. Choi et al. (2012) compared the performance of PROVEAN with the results from two different protein databases. This included the NCBI NR (non-redundant) protein and the UniProtKB/Swiss-Prot protein databases. Their results indicated a reduced accuracy of 7% when using the UniProtKB/Swiss-Prot database instead of the NCBI NR protein database. The authors highlight the usefulness of the program to identify deleterious single nucleotide variants and variants that cause protein sequence indels. We found this program useful for predicting deleterious SNPs consistent with other prediction servers.

To predict the pathology of the identified 14 nsSNPs, the PMut program was utilized. Of 86 nsSNPs tested, 37 including the 14 nsSNPs were predicted to be associated with a pathological disease. PMut is reported to be a powerful tool to predict the functional consequences of protein sequence variants (Lopez-Ferrando et al. 2017).

We utilized six different bioinformatics programs (SIFT, PolyPhen, PANTHER, PROVEAN, Mutation Assessor and P-Mut) that use different methods to predict the nsSNP with deleterious effect on the *GABRA2* protein function. Fourteen, out of 86 nsSNPs that were predicted as most damaging in all 6 programs that we used were further analyzed for structural stability changes.

Among the 14 nsSNPs, 10 were found in the conserved regions of the protein sequence as identified by ConSurf analysis. The nsSNPs that are located at highly conserved amino acid positions tend to be more deleterious than nsSNPs that are located at non-conserved sites. In general, highly conserved amino acids either buried (structural) or exposed (functional) act as biologically active sites compared to other residues. Any substitutions in these functional residues may either lead to complete loss of biological functions or cause severe deleterious effects compared to other polymorphisms of the non-conserved site (Dakal et al. 2017).

The functional and structural sites were identified in ConSurf program that combines evolutionary data and solvent accessibility predictions. In our study, among the 14 nsSNPs, 6 nsSNPs (P445H, P280S, V208F, S186C, T43A, R58T) were found in the exposed surface and 8 (I255T, L170S, W122L, I121N, F93C, Y73C, F93C, A188T) were buried.

We, therefore, analyzed the predicted structural consequences using tools available in Swiss-PDB viewer. The 3D structures of variants and the wild type were generated in I-TASSER program. The model that had a high C-score for each of the variants and the wild type was selected for analysis in the DeepView Swiss-PDBViewer.

The total energy of the native protein structure was determined to be -19946 KJ/mol. Among all the 14 mutants, the mutant R58T was found to have the highest energy of -19646 KJ/mol after energy minimization. The remaining 13 mutants were found to have

much lower energy values ranging from - 13567.015 to -13857.964. The variant that had a high difference in RMSD score of 0.150Å is S186C. The Ramachandran plot as analyzed in RAMPAGE program indicated no major changes in terms of shifts to the favoured region or allowed regions. The number of residues in both regions remained the same. However, the I-Mutant suite which was used to analyze the protein structural stability changes predicted 10 of 14 nsSNPs as disease-related and 3 nsSNPs as neutral polymorphisms. The reliability index ranged from 2 to 10 for the 13 variants.

In our study, of 14 variants analyzed, one variant, R58T had DDG value of 0.23 indicating a weak effect and 12 other variants (excluding I255T) had scores less than - 0.5 indicating largely destabilizing effect and two variants, W122L and S186C with near scores of -0.49 and - 0.47 respectively indicating weak effect. Interestingly, among the 14 variants analyzed, 13 except P445H were present in the neurotransmitter-gated ion-channel ligand-binding domain of the protein. This indicates the potential role of the SNPs in the functionality of the protein.

While studying alcohol abuse, we have to be conscious of "Emergent Complexity" *i.e.* alcoholism could be a product of multi-factorial elements contributing to this psychiatric condition. One of several such elements would be the polymorphism of important proteins which contribute to change in function of pathways in the central nervous system. One such observation has been the SNPs in GABRA2 gene. Our study showed the effect of the SNPs in the structure and function of the protein (Agrawal et al. 2012).

The MutPred program predicts the impact of single amino acid substitutions on more than 50 different protein properties to infer the molecular mechanisms of pathogenicity. The software package includes genetic and molecular data of amino acid substitutions leading to varied pathology. It includes a general pathological prediction and a ranked list of specific molecular alterations potentially affecting the phenotype (Pejaver et al. 2017).

In our study, two variants (W122L, I121N) were available in the dbSNP database with a frequency of < 0.01. Nine variants (R58T, Y73C, W122L, L170S, I255T, A188T, P280S, P426S, P445H) were available and were identified as "rare" (Minor Allele Frequency < 0.01) variations by the Exome Aggregation Consortium (ExAC) database (<http://exac.broadinstitute.org/>). This database lists a total of 53 missense variants with a constraint metric (z value) of 3.34. Positive Z scores indicate increased constraint (intolerance to variation) and therefore that the gene had fewer variants than expected.

The rare functional variant could alter gene function significantly though hit occurs at low frequency in a population. The "common-disease rare-variant" hypothesis indicates that variants affecting health are under purifying selection and thus should be found only at low frequencies in human populations. Rare variants are increasingly being studied, as a consequence of exome and whole-genome sequencing efforts. While these variants are individually infrequent in populations, there are many such variants in human populations, and they can be unique to specific populations. They are more likely to be deleterious than common variants, as a result of rapid population growth and weak purifying selection (Nelson et al. 2012).

Our overall results indicate that there is a significant number of nsSNP (14/89; 16%) predicted to be associated with GABRA2 protein dysfunction. Gene-gene interactions were studied to highlight candidate genes that could be associated with alcohol dependence, especially if haplotypes are to be studied in the future. Among the 14 nsSNPs, 9 were within conserved regions. The functionally deleterious nsSNPs, showed 10 nsSNPs to be associated with disease, though none of them showed structural variation in the Ramachandran Plot.

GABRA2 gene variants are associated with alcohol dependence and other mental disorders, but nsSNP of the GABRA2 gene has not been studied earlier. To our knowledge, this is the first report on the SNPs focusing primarily on exons associated with functional changes directed by missense variants in the gene. It could be speculated that these nsSNPs play a vital role in an individual's alcohol dependence. However, this should be further evaluated through studies on individuals with varying degrees of alcohol dependence in addition to Genome-Wide Association Studies (GWAS). A case-control study involving alcoholics compared to non-alcoholics with SNPs will help rule out type I error in observations (Ray et al. 2009). This data could help institute social intervention programs. With a better understanding of the genetic basis of alcoholism, it is possible to pre-screen 'at-risk' individuals and design personalized early intervention, especially among the youth population (counselling, change of peer groups, improve family support and ensure employment and domiciliary status if necessary).

Declarations

Conflicts of Interest:

None

References

1. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–9.
2. Agrawal A, Edenberg HJ, Foroud T, Bierut LJ, Dunne G, Hinrichs AL, Nurnberger JI, Crowe R, Kuperman S, Schuckit MA, Begleiter H, Porjesz B, Dick DM. Association of GABRA2 with drug dependence in the collaborative study of the genetics of alcoholism sample. *Behav Genet*. 2006;36:640–50.
3. Agrawal A, Verweij KJ, Gillespie NA, Heath AC, Lessov-Schlaggar CN, Martin NG, Nelson EC, Slutske WS, Whitfield JB, Lynskey MT. The genetics of addiction—a translational perspective. *Transl Psychiatry*. 2012 Jul 17;2:e140.
4. Begleiter H, Porjesz B. Genetics of human brain oscillations. *Int J Psychophysiol Off J Int Organ Psychophysiol*. 2006;60:162–71.
5. Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Grucza RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman S, Lynskey M, Mann K, Neuman RJ, Nöthen MM, Nurnberger JI Jr, Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP; Gene, Environment Association Studies Consortium. A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A*. 2010;107:5082–7.
6. Birkley EL, Smith GT. Recent advances in understanding the personality underpinnings of impulsive behavior and their role in risk for addictive behaviors. *Curr Drug Abuse Rev*. 2011;4:215–27.
7. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLOS ONE*. 2012;7(10):e46688.
8. Clarke T-K, Adams M J, Davies G, Howard D M, Hall L S, Padmanabhan S, Murray A D, Smith BH, Campbell A, Hayward C, Porteous DJ, Deary IJ, McIntosh AM. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Mol Psychiatry*. 2017; 22: 1376–1384
9. Covault J, Gelernter J, Hesselbrock V, Nellissery M, Kranzler HR. Allelic and haplotypic association of GABRA2 with alcohol dependence. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet*. 2004;129B:104–9.
10. Dakal TC, Kala D, Dhiman G, Yadav V, Krokhotin A, Dokholyan NV. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. *Sci Rep*. 2017;7:6525.
11. Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, Crowe RR, Goate A, Hesselbrock V, Jones K, Kwon J, Li TK, Nurnberger JI Jr, O'Connor SJ, Reich T, Rice J, Schuckit MA, Porjesz B, Foroud T, Begleiter H. Variations in GABRA2, Encoding the $\alpha 2$ Subunit of the GABAA Receptor, Are Associated with Alcohol Dependence and with Brain Oscillations. *Am J Hum Genet*. 2004;74:705–14.
12. Farren CK, Hill KP, Weiss RD. Bipolar Disorder and Alcohol Use Disorder: A review. *Curr Psychiatry Rep*. 2012;14:659–66.
13. Feldstein Ewing SW, LaChance HA, Bryan A, Hutchison KE. Do genetic and individual risk factors moderate the efficacy of motivational enhancement therapy? Drinking outcomes with an emerging adult sample. *Addict Biol*. 2009;14:356–65.
14. Gonzalez-Nunez V. Role of gabra2, GABA receptor alpha-2 subunit, in CNS development. *Biochem Biophys Rep*. 2015;3:190–201.
15. Hassan MM, Omer SE, Khalfallah RM, Mustafa RY, Ali IS, Mohamed SB. Bioinformatics Approach for Prediction of Functional Coding/Noncoding Simple Polymorphisms (SNPs/Indels) in Human BRAF Gene. *Adv Bioinforma* 2016;2016:2632917.
16. Huang YH, Liu HC, Tsai FJ, Sun FJ, Huang KY, Chiu YC, Huang YH, Huang YP, Liu SI. Correlation of impulsivity with self-harm and suicidal attempt: a community study of adolescents in Taiwan. *BMJ Open*. 2017;7:e017949.
17. Littlefield AK, Sher KJ, Wood PK. A personality-based description of maturing out of alcohol problems: extension with a five-factor model and robustness to modeling challenges. *Addict Behav*. 2010;35:948–54.

18. López-Ferrando V, Gazzo A, de la Cruz X, Orozco M, Gelpí JL. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res.* 2017;45(W1):W222–8.
19. Mayfield RD, Harris RA, Schuckit MA. Genetic factors influencing alcohol dependence. *Br J Pharmacol.* 2008;154:275–87.
20. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu SA, Fraser D, Warren L, Aponte J, Zawistowski M, Liu X, Zhang H, Zhang Y, Li J, Li Y, Li L, Woollard P, Topp S, Hall MD, Nangle K, Wang J, Abecasis G, Cardon LR, Zöllner S, Whittaker JC, Chisoe SL, Novembre J, Mooser V. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science.* 2012;337(6090):100–4. s
21. Nusbaumer MR, Reiling DM. Environmental influences on alcohol consumption practices of alcoholic beverage servers. *Am J Drug Alcohol Abuse.* 2002;28:733–42.
22. Pejaver V, Mooney SD, Radivojac P. Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. *Hum Mutat.* 2017;38:1092–108.
23. Ray LA, Hutchison KE. Associations among GABRG1, level of response to alcohol, and drinking behaviors. *Alcohol Clin Exp Res.* 2009 Aug;33(8):1382-90.
24. Schaefer C, Rost B. Predict impact of single amino acid change upon protein structure. *BMC Genomics.* 2012;13(Suppl 4):S4.
25. Soyka M, Preuss UW, Hesselbrock V, Zill P, Koller G, Bondy B. GABA-A2 receptor subunit gene (GABRA2) polymorphisms and risk for alcohol dependence. *J Psychiatr Res.* 2008;42:184–91.
26. Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinforma Oxf Engl.* 2016;32:2230–2.
27. Taylor EM, Murphy A, Boyapati V, Ersche KD, Flechais R, Kuchibatla S, McGonigle J, Metastasio A, Nestor L, Orban C, Pasetti F, Paterson L, Smith D, Suckling J, Tait R, Lingford-Hughes AR, Robbins TW, Nutt DJ, Deakin JF, Elliott R; ICCAM Platform. Impulsivity in abstinent alcohol and polydrug dependence: a multidimensional approach. *Psychopharmacology (Berl).* 2016;233:1487–99.
28. Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, Fehr C, Scherbaum N, Steffens M, Ludwig KU, Frank J, Wichmann HE, Schreiber S, Dragano N, Sommer WH, Leonardi-Essmann F, Lourdasamy A, Gebicke-Haerter P, Wienker TF, Sullivan PF, Nöthen MM, Kiefer F, Spanagel R, Mann K, Rietschel M. Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry.* 2009;66:773–84.
29. Villafuerte S, Strumba V, Stoltenberg SF, Zucker RA, Burmeister M. Impulsiveness mediates the association between GABRA2 SNPs and lifetime alcohol problems. *Genes Brain Behav.* 2013;12:525–31.
30. Way MJ, Ali MA, McQuillin A, Morgan MY. Genetic variants in ALDH1B1 and alcohol dependence risk in a British and Irish population: A bioinformatic and genetic study. *PloS One.* 2017;12:e0177009.

Tables

Table 1

Analysis of the 14 nsSNPs identified in terms of nature of mutation, energy minimization and structural integrity of the wild and mutation-predicted model.

AA variant	Variant ID	I-MUTANT			Swiss-PDBViewer	RAMPAGE (No. of residues)			
		I-Mutant Disease	RI	DDG Value (kcal/mol)	Total energy after energy minimization (KJ/mol)	RMSD (Å)	Favored	Allowed	Outlier
Wild	-	-	-	-	-19946.123	0.00	339 (75.5%)	67 (14.9%)	43 (9.6%)
T43A	rs41305781	Disease	8	-1.33	-13702.660	0.000	339 (75.5%)	67 (14.9%)	43 (9.6%)
R58T	rs765251624	Disease	3	0.23	-19646.230	0.000	339 (75.5%)	67 (14.9%)	43 (9.6%)
Y73C	rs753040126	Disease	5	-0.98	-13652.485	0.000	339 (75.5%)	67 (14.9%)	43 (9.6%)
F93C	rs199725032	Disease	7	-2.34	-13680.411	0.029	339 (75.5%)	67 (14.9%)	43 (9.6%)
I121N	rs749035438	Disease	6	-0.57	-13857.964	0.059	339 (75.5%)	67 (14.9%)	43 (9.6%)
W122L	rs775541780	Disease	8	-0.49	-13567.015	0.081	339 (75.5%)	67 (14.9%)	43 (9.6%)
S186C	rs373038663	Disease	4	-0.47	-13630.729	0.150	339 (75.5%)	67 (14.9%)	43 (9.6%)
A188T	Rs768736908	Disease	8	-1.05	-13724.141	0.023	339 (75.5%)	67 (14.9%)	43 (9.6%)
V208F	rs752066816	Neutral	2	-0.46	-13739.968	0.033	339 (75.5%)	67 (14.9%)	43 (9.6%)
I255T	rs767460850	Unknown	-	-	-13738.369	0.055	339 (75.5%)	67 (14.9%)	43 (9.6%)
P280S	rs771481457	Disease	8	-2.01	-13763.594	0.057	340 (75.7%)	66 (14.7%)	43 (9.6%)
P426S	rs754301188	Neutral	9	-0.86	-13749.602	0.067	339 (75.5%)	67 (14.9%)	43 (9.6%)
P445H	rs761715134	Disease	10	-2.48	-13763.185	0.065	339 (75.5%)	67 (14.9%)	43 (9.6%)
L170S	rs749473765	Neutral	7	-0.74	-13684.243	0.047	339 (75.5%)	67 (14.9%)	43 (9.6%)
<p>RI - reliability index The RI value (Reliability Index) is computed only when the sign of the stability change is predicted DDG Value - Free energy change value The DDG value is calculated from the unfolding Gibbs free energy value of the mutated protein minus the unfolding Gibbs free energy value of the wild type (Kcal/mol). RMSD - Root Mean Square Deviation</p>									

Table 2

Molecular changes and prediction by Mutpred server

AA variation	Molecular mechanisms	g-score	P-score
T43A	Altered Transmembrane protein***	0.874	2.2e-03
	Gain of N-linked glycosylation at N38***		6.7e-03
R58T		0.945	
	Loss of Allosteric site at R58***		1.8e-04
	Altered Metal binding***		3.8e-03
	Loss of Relative solvent accessibility**		0.01
	Altered Disordered interface**		0.04
	Altered Ordered interface**		0.01
	Altered Transmembrane protein***		2.0e-03
	Altered DNA binding**		0.01
	Loss of Catalytic site at R58**		0.04
Gain of Sulfation at Y53**	0.04		
Y73C	Altered Disordered interface***	0.895	2.0e-03
	Altered Ordered interface***		1.6e-03
	Altered Transmembrane protein***		2.7e-03
F93C	Altered Ordered interface***	0.936	9.5e-03
	Altered Metal binding**		0.02
	Altered DNA binding**		0.01
	Gain of Allosteric site at D90*		0.05
	Altered Transmembrane protein**		0.01
	Gain of Pyrrolidone carboxylic acid at Q95**		0.03
I121N	Loss of Allosteric site at W122***	0.954	7.5e-03
	Altered Ordered interface**		0.04
	Altered Transmembrane protein***		2.7e-03
	Altered Metal binding**		0.01
	Gain of Ubiquitylation at K120**		0.04
	Altered Stability**		0.03
W122L	Loss of Allosteric site at W122***	0.937	1.5e-03
	Altered Ordered interface***		5.3e-03
	Loss of Strand**		0.02
	Altered Metal binding**		0.02

	Altered Transmembrane protein***		8.6e-03
	Gain of Ubiquitylation at K120**		0.04
L170S	Gain of Intrinsic disorder**		0.01
	Altered Ordered interface**	0.935	0.01
	Altered Metal binding***		5.2e-03
	Altered Transmembrane protein***		9.6e-04
S186C	Altered Transmembrane protein***	0.855	5.4e-05
	Altered Ordered interface***		8.2e-03
A188T	Altered Transmembrane protein**		2.9e-05
	Loss of Relative solvent accessibility*	0.745	0.03
	Altered Stability*		0.02
V208F	Altered Transmembrane protein***	0.896	2.9e-05
	Loss of Relative solvent accessibility**		0.03
	Altered Stability**		0.02
I255T	Altered Transmembrane protein***	0.880	9.7e-06
	Altered Ordered interface**		0.02
	Altered Stability**		0.04
P280S	Altered Transmembrane protein***		4.2e-04
	Gain of Relative solvent accessibility**	0.899	0.03
	Altered Metal binding**		0.04
P426S	Altered Ordered interface**		0.05
	Loss of Allosteric site at R422**	0.797	0.02
	Altered Transmembrane protein***		3.9e-03
	Altered DNA binding**		0.02
P445H	Altered Metal binding***		3.3e-03
	Altered Ordered interface***	0.825	4.8e-03
	Loss of Loop**		0.04
	Altered Transmembrane protein***		1.2e-03
	Loss of Allosteric site at Y440**		0.03

Certain combinations of high values of general scores and low values of property scores are referred to as hypotheses

Scores with g -value > 0.5 and p -value < 0.05 are referred to as actionable hypotheses (*)

Scores with g -value > 0.75 and p -value < 0.05 are referred to as confident hypotheses (**)

Scores with g -value > 0.75 and p -value < 0.01 are referred to as very confident hypotheses. (***)

The output of MutPred contains a general score (g), i.e., the probability that the amino acid substitution is deleterious/disease-associated, and top 5 property scores (p), where p is the P-value that certain structural and functional properties are impacted.

Figures

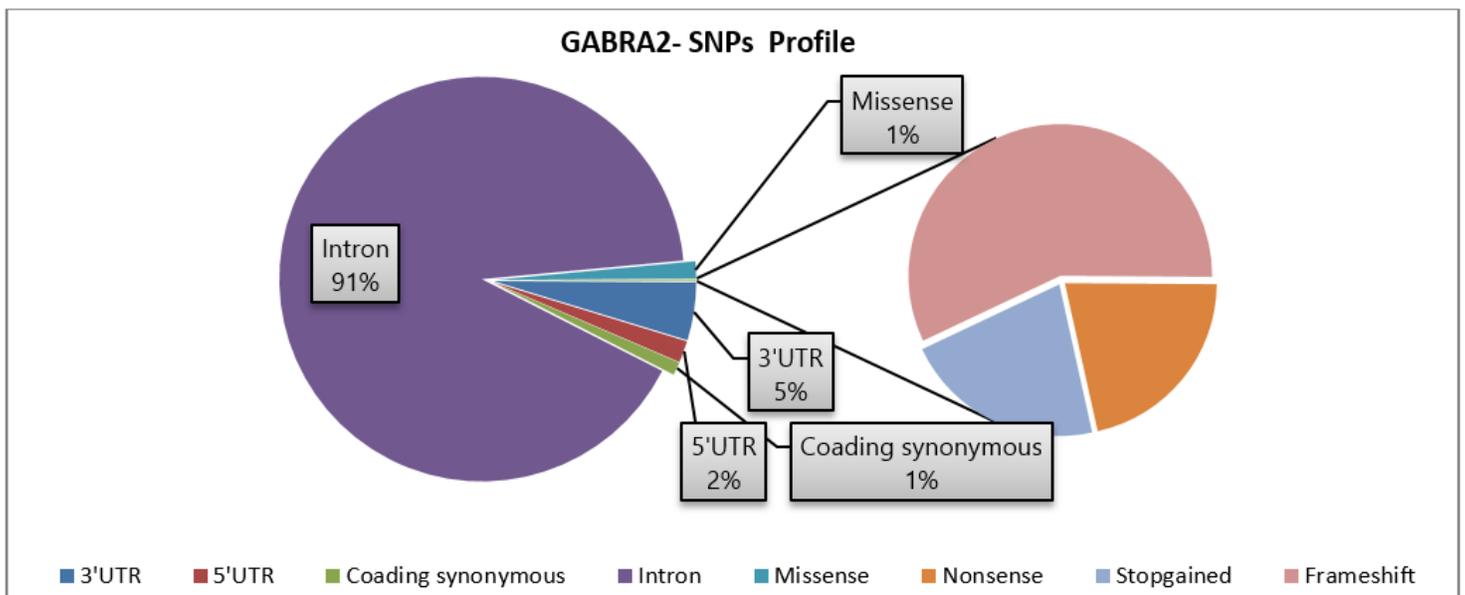


Figure 1

Proportional distribution of SNPs of GABRA2 gene belonging to different functional classes and untranslated regions

The GABRA2 (isoform 1) (accession code: P47869 GABRA2_HUMAN) was downloaded from Uniprot (3). Please note the missense SNPs are a minority which was analyzed. The derived second pie chart indicates proportions of nonsense, stopgained and frameshift mutations.

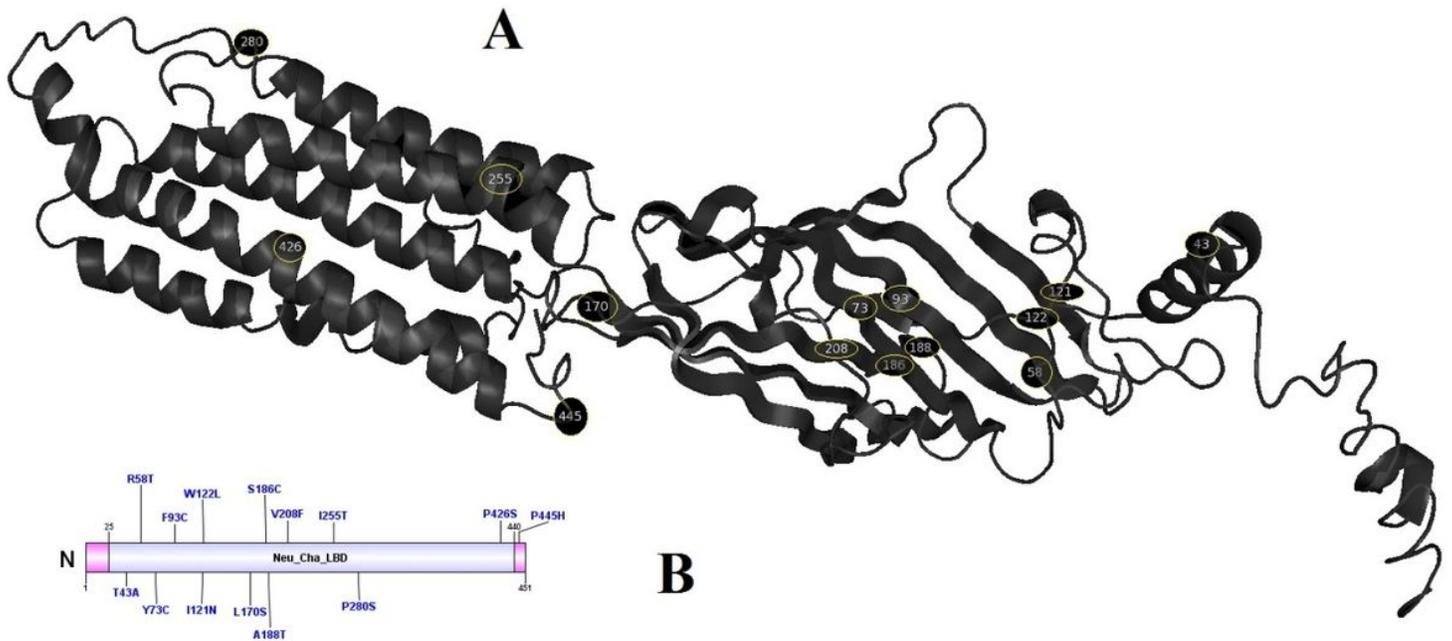


Figure 2

3D structure of native GABRA2 protein indicating the identified 14 deleterious SNPs (A) and 13 SNP positions in the neurotransmitter-gated ion-channel ligand-binding domain (B)

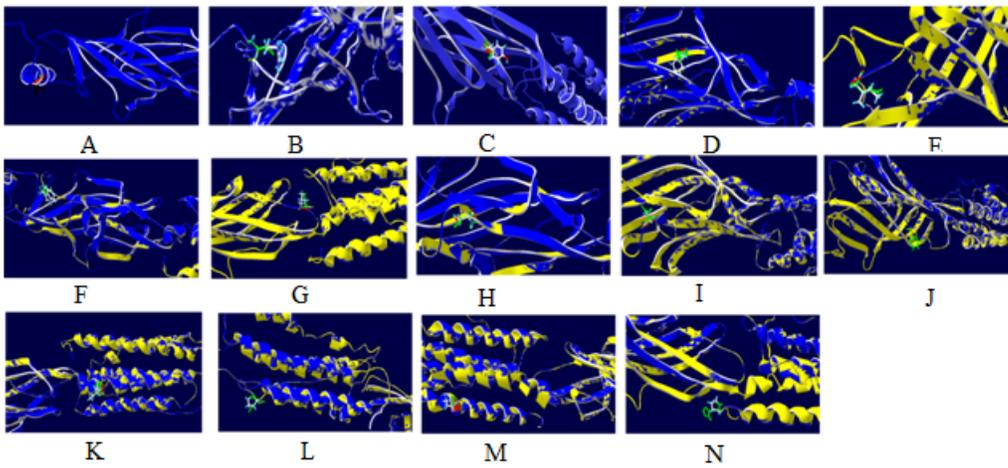


Figure 3

Superimposed mutant protein structures of GABRA2 protein

The 3D models of the variants P445H, P280S, I255T, V208F, S186C, L170S, W122L, I121N, F93C, Y73C, R58T and T43A are labeled from A through N respectively.

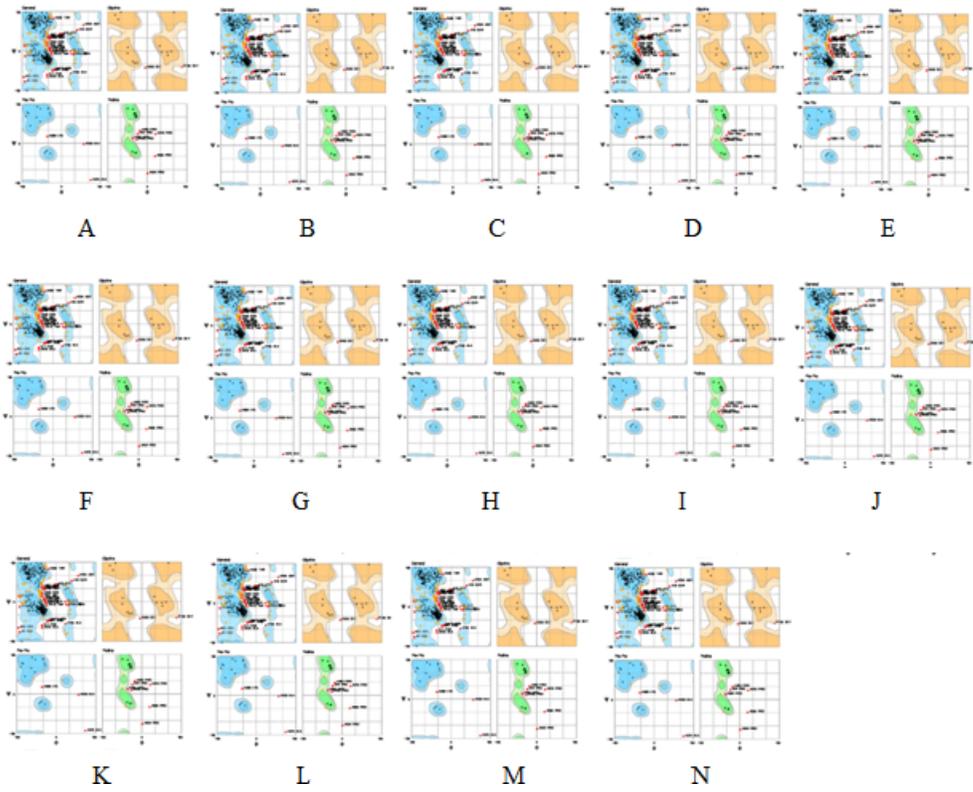


Figure 5

Ramachandran plot for mutants

The Ramachandran plots of the variants P445H, P280S, I255T, V208F, S186C, L170S, W122L, I121N, F93C, Y73C, R58T and T43A are labeled from A through N respectively.



The conservation scale:
 1 2 3 4 5 6 7 8 9
 Variable Average Conserved

Figure 6

Consurf analysis of human GABRA2 protein

- f- A predicted functional residue (highly conserved and exposed).
- s- A predicted structural residue (highly conserved and buried). The mutants' location are indicated in red arrows.

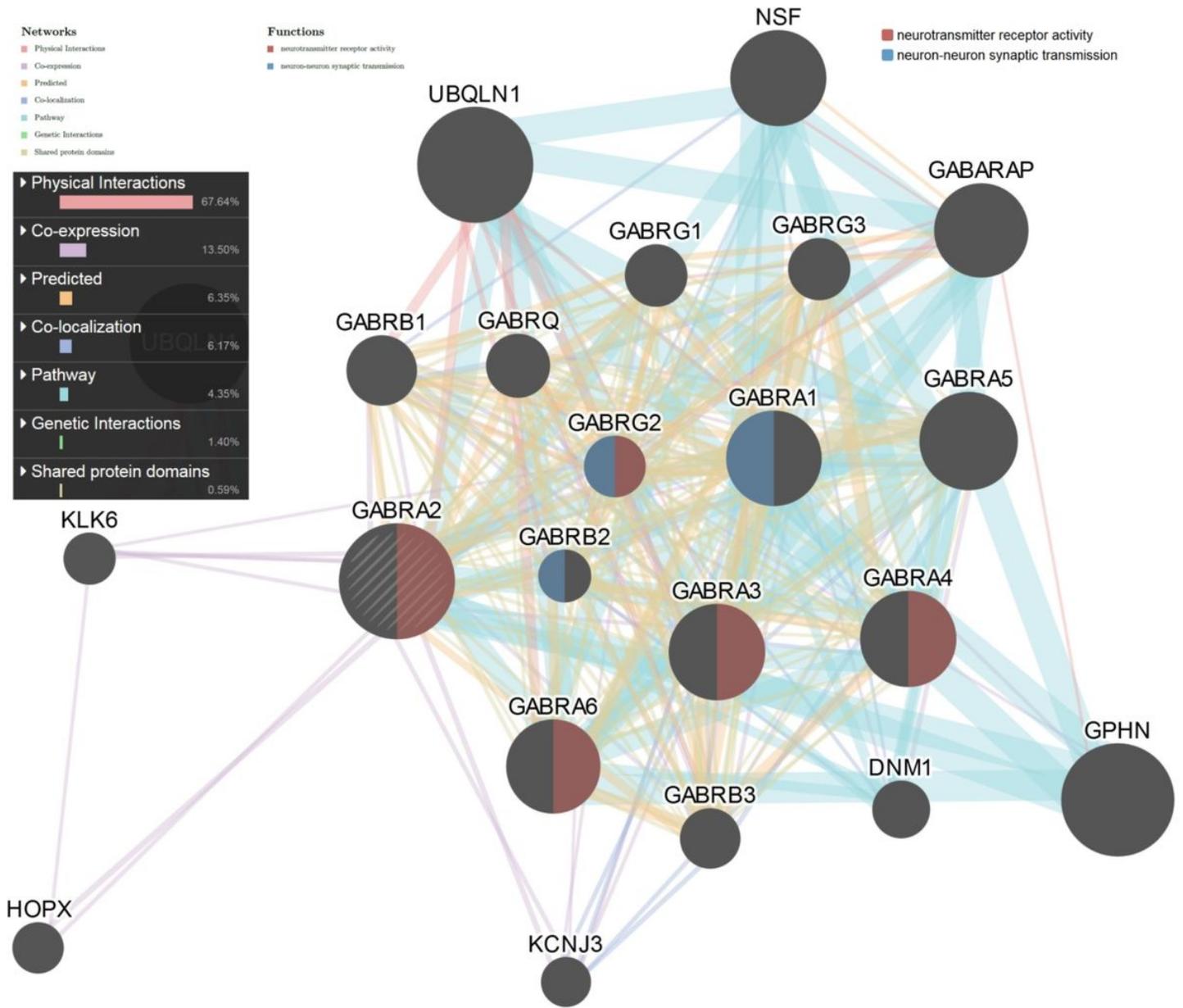


Figure 7

Protein-Protein interaction network of GABRA2 protein predicted by GENEMANIA program