International Journal of Current Research and Review
DOI: http://dx.doi.org/10.31782/IJCRR.2020.122030

Scopus

Research Article

IJCRR
Section: Healthcare
Sci. Journal Impact Factor: 6.1 (2018)
ICV: 90.90 (2018)

In-*silico* Analysis of Deleterious Single Nucleotide Polymorphisms (SNPs) and Molecular Docking of Disease-linked Mutations in Genes Responsible for Schizophrenia

Neema Tufchi[1], Kumud Pant[1], Devvret[1], Akshara Pandey[2]

[1]Department of Biotechnology, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India; [2]Department of Computer Application, Graphic Era Hill University, Dehradun, Uttarakhand, India.

ABSTRACT

Introduction: Schizophrenia (SZ) is a neurological disorder, the causative agents of which may be multiple factors like genetic, environmental factors, or co-morbidities with other diseases. The actual reason for the occurrence of this disorder is yet to be unrevealed. The genes responsible for this disorder are vulnerable to mutations at the chromosomal or protein levels. So identification of disease-associated mutations may pave the way for divulging the root cause behind the disorder. In the current study, the emphasis had been made on finding the said disease-associated mutations for the disorder through *in-silico* analyses.

Methods: The genes and FDA approved antipsychotics were prioritized using text mining approach, which shortlists nine genes (*COMT, DISC1, DAOA, NRG1, PRODH, RGS4, GRM3, DRD3* and *DTNBP1*) and seven antipsychotics (Haloperidol, Fluphenazine, Aripiprazole, Clozapine, Iloperidone, Lurasidone, and Risperidone). The genes were checked for deleterious or damaging mutations using SIFT and PolyPhen servers.

Results: The SNPs rs6267 and rs4986871in COMT protein were found deleterious with both the servers. SNPs rs2391191 and rs9558562 were found damaging in DAOA protein. In case of DISC1 protein five SNPs (rs6675281, rs821616, rs3738400, rs34622148, and rs55795950) were found damaging. NRG1 and RGS4 protein have one deleterious mutation (rs3924999 and rs68678746 respectively) and three deleterious mutations (rs450046, rs2870984 and rs397055) were present in PRODH protein. The SNPs rs181422088 (in DRD3) and rs16876589 (in DTNBP1) were found deleterious with both the servers. The native protein and their mutated form were modelled and docked with the antipsychotics to check their binding energies.

Conclusion: The results showed that the binding energies between antipsychotics and mutated proteins were lower as compared to native protein suggesting that mutated proteins bind well and were stable, so a person is prescribed antipsychotics to reduce the symptoms of the disorder. Thus, these mutations may be the reason behind the pathophysiology of the disorder.

Key Words: Antipsychotics, Docking, Genes, *In-silico*, Mutations, Schizophrenia.

INTRODUCTION

In the current study, in-silico methods and servers were used to predict the mutations in the gene candidate which may have a role in the cause of schizophrenia (SZ). *In-silico* methods/approaches have set foot in modern pharmacology, drug identification, and discovery. *In-silico* methods have several advantages including fast predictions, safety and time, and cost-effectiveness [1]

SZ is a stern and chronic disorder, affecting the thinking and social behaviour of an individual. People affected with this disorder seem like they have lost touch with reality i.e., they live in their world of hallucinations and delusions [2]. The disorder is not as common as other psychiatric disorders, and the symptoms are very disabling. According to WHO status 2018, schizophrenia has affected 20 million people in the world, men in their early age from 13 to 25 years, and women of age 25-35 years being most susceptible [2]. The disorder may affect the occupational and educational behaviour of an individual. People with schizophrenia are more vulnerable to die 2-3 times early as compared to the general population because of the combination of many diseases like metabolic, infectious, and cardiovascular diseases [3].

SZ is a complex disorder with unknown etiology, with ap-

Corresponding Author:

Neema Tufchi, Department of Biotechnology, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India.
Email: neematufchi@gmail.com

ISSN: 2231-2196 (Print) **ISSN:** 0975-5241 (Online)

Received: 24.07.2020 **Revised:** 22.08.2020 **Accepted:** 18.09.2020 **Published:** 27.10.2020

proximately 800 genes (identified through GWAS) which may have a role in the susceptibility of SZ [4]. In many studies, it has been found that mutations in the genes are responsible for the pathophysiology of the disorder [5]. The mutations may arise during DNA replication or due to environmental factors. For about half of the human inherited diseases, amino acid substitutions are responsible [6]. The SNPs (single nucleotide polymorphism) identification is significant for predicting an individual's risk of developing diseases, response to certain drugs and susceptibility to environmental factors such as toxins [7]. The SNPs are categorized into non-synonymous (nsSNPs) and synonymous mutations.

The people diagnosed with SZ are treated with antipsychotics as prescribed by neurologists. There are more than 20 FDA approved antipsychotics for neurological disorders. The drugs are broadly categorized into first-generation or typical drugs and second-generation or atypical drugs [8,9]. The genes and antipsychotics studied are briefly described in tables 1 and 2 respectively.

MATERIALS AND METHODS

Gene retrieval and prioritization

The information regarding genes responsible for the cause of schizophrenia was obtained using text mining. The text mining approaches include information retrieval, clustering, document classification, identification of data trends which can be used for ranking of genes. Keywords used for extracting data from the scientific literature are very crucial for gene prioritization. For finding the genes responsible for causing schizophrenia PubTator was used, with the search terms "schizophrenia genes", "schizophrenia genes SNPs" and "schizophrenia genes mutations SNPs". After shortlisting with PubTator nine prioritized genes *COMT, DISC1, DAOA, NRG1, PRODH, RGS4, GRM3*, *DTNBP1* and DRD3 were chosen for further analyses.

The nsSNPs of two proteins were retrieved from the dbSNP database. The SNPs obtained from dbSNP were then analyzed using PolyPhen and SIFT servers [10].

Analysis of nsSNPs by SIFT

Sort Intolerant From Tolerant (SIFT) is a tool that distinguishes between tolerant and intolerant amino acid substitutions. The tool is based on the theory that protein function is related to protein phylogeny. The functionally important residues remain conserved in the sequence while insignificant residues vary in the alignment. When protein sequence is submitted to SIFT, it sorts the intolerant amino acid from tolerant by generating a score output file in which intolerant residues are highlighted in red. If a query contains single amino acid substitutions the tool can predict mutants having phenotypic effects [11].

Analysis of ns SNPs by PolyPhen

Polymorphism Phenotyping (PolyPhen) works on a Bayesian approach. It predicts the effect of variants by using both structure and sequence information. For the identification of functional annotation it creates clustered and multiple sequence alignment. The tool also predicts identity and profile-based scores with structural properties like B factor, solvent accessibility, and hydrophobic propensity, etc. All the properties are combined using two Bayesian models, HumVar and HumDiv. These two probabilistic models were trained on different datasets. HumVar was trained based on differences among disease-causing mutations in the UniProtKB database and nsSNPs with no disease phenotype. HumVar model is generally ideal for Mendelian diseases because it distinguishes the diseased mutation from normal human variants. HumDiv distinguishes damaging alleles (having known effect on function) from non-damaging alleles. HumDiv model identifies variants with slightly deleterious alleles and treats them as damaging. PolyPhen determines prediction and threshold by calculating true positive and false-positive rates and thus predicts the "probably damaging", "possibly damaging" and benign residues [12].

Modelling of nine proteins and their mutant sequences

Native and mutant protein structures were analyzed to study their stability. The native structure of PRODH, DTNBP1, DISC1 were not present in the PDB database and other proteins like COMT, NRG1, GRM3, DAOA, RGS4 and DRD3 were present in the database but in ligand bounded form. Ab-initio protein modelling was performed to determine the three-dimensional structure of the native protein (PRODH, DTNBP1 and DISC1) using the I-Tasser server. Mutated residues were predicted after analyzing with SIFT and PolyPhen. Mutated structures were modelled and energy minimization was done by KoBaMIN server [13]. RMSD (Root mean square deviation) values were studied for the deviation of mutant structure from the native structure [14].

Three-dimensional structures of other proteins (COMT, NRG1, GRM3, DAOA, RGS4 and DRD3) were present in the database but in the bounded form with ligands, so to analyze the structure of the only protein, the ligand was removed from the structure using Pymol [15].

Screening of antipsychotic drugs

The drugs were screened as per prescription by neurologists for lowering down the symptoms of schizophrenia. The screened drugs were treated as ligands. The FDA approved drugs were screened using a text mining approach which prioritizes the drugs according to their effect on the disorder. PubTator was used for the text mining approach with keywords "schizophrenia", "antipsychotic drugs" and "FDA approved". After applying these filters the seven FDA

approved drugs (Haloperidol, Fluphenazine, Aripiprazole, Clozapine, Iloperidone, Lurasidone, and Risperidone) were selected for in-silico docking purpose. The structures were downloaded from an open-source PubChem database. The canonical SMILES of these compounds were downloaded from PubChem and converted into .pdb format using Open Babel which is format inter-conversion software [16].

Docking Studies

AutodockVina in PyRx was used for the molecular docking studies of nine proteins and their mutated forms as receptors. The ligands were Haloperidol, Fluphenazine, Aripiprazole, Clozapine, Iloperidone, Lurasidone, and Risperidone. The .pdb format of receptor and ligand is uploaded as an input file, which gets converted into PDBQT format.

RESULTS AND DISCUSSION

A total of nine genes were shortlisted because of their role in the pathophysiology of schizophrenia and the number of cited literature of these genes was very high as compared to other genes (Figure 1). The number of studies on COMT and schizophrenia is shown as 655 (curated on 15-02-2020) on PubTator by using keywords "COMT, Schizophrenia".

There are 510 studies showing association of NRG1 gene and schizophrenia in PubTator using keywords "NRG1, Schizophrenia". In the case of GRM3, there are 78 published literature and the keywords used were "GRM3, Schizophrenia". DAOA and PRODH have 145 and 61 literature respectively and keywords used were "DAOA, Schizophrenia" and "Proline dehydrogenase gene, schizophrenia" respectively. In case of DTNBP1, DISC1, RGS4 and DRD3 the result comes to be of 314, 700, 113, and 310 respectively.

SNP dataset of human proteins from dbSNP

SNPs in the studied genes viz. COMT, NRG1, RGS4, DISC1, DTNBP1, GRM3, DRD3, DAOA and PRODH were retrieved from dbSNP database. In the case of COMT, a total of 8235 SNPs were present which include 298 coding non-synonymous SNPs and 336 non-coding SNPs. 541 non-coding SNPs were located in 3' UTR region and 320 SNPs in 5' UTR region. The rest of the SNPs were distributed in the intron region (7791), stop gained region (19), synonymous coding region (177) etc. For NRG1 a total of 260749 SNPs were present that includes 892 coding non-synonymous SNPs and 2195 non-coding SNPs. Among non-coding SNPs, 3053 were present in 3' UTR region and 1198 were located in 5' UTR region. In RGS4 2663 SNPs were present having 203 coding non-synonymous SNPs and non-coding SNPs were not available for this gene. Total 549 non-coding SNPs were located in 3' UTR region and 37 SNPs in 5' UTR region. For DISC1 a total of 93028 SNPs were present included 786 coding non-synonymous SNPs and 4384 non-coding SNPs. Among non-coding SNPs, 2587 were present in 3' UTR region and 37 were located in 5' UTR region. In the case of DTNBP1, a total of 31557 SNPs were present which include 342 coding non-synonymous SNPs and 2010 non-coding SNPs. 480 non-coding SNPs were located in 3' UTR region and 364 SNPs in 5' UTR region. In GRM3, 50136 SNPs were present having 777 coding non-synonymous SNPs and 2395 non-coding SNPs. 509 non-coding SNPs were located in 3' UTR region and 301 SNPs in 5' UTR region. In DRD3, 16366 SNPs were present having 250 coding non-synonymous SNPs and 440 non-coding SNPs. Overall 205 non-coding SNPs were located in 3' UTR region and 235 SNPs in 5' UTR region. In the case of DAOA, a total of 7208 SNPs were present which include 249 coding non-synonymous SNPs and 192 non-coding SNPs. 119 non-coding SNPs were located in 3' UTR region and 205 SNPs in 5' UTR region. In PRODH, 5329 SNPs were present having 450 coding non-synonymous SNPs and 540 non-coding SNPs. 478 non-coding SNPs were located in 3' UTR region and 62 SNPs in 5' UTR region. In the present study, nsSNPs (coding non-synonymous SNPs) and non-coding SNPs in 3' and 5' UTR regions were studied.

Damaging nsSNPs predicted by SIFT program

SIFT was used to study nsSNPs, the nsSNPs predicted from genes were submitted to SIFT for the analysis of damaging mutations or tolerance indices. Lesser the functional impact, higher is the tolerance index of amino acid and vice versa. The results of SIFT for nine genes are shown in table 3.

Total 536 nsSNPs for COMT were uploaded to SIFT for the tolerance index analysis, 17 SNPs were found to be damaging or deleterious having tolerance index of ≤ 0.005. 202 nsSNPs of DAOA were submitted to SIFT out of which 23 were found deleterious each having tolerance index of 0.00. In the case of DISC1, 1091 nsSNPs were submitted to SIFT, 28 of which were found to be damaging/deleterious with tolerance index ≤ 0.005. Only 20 nsSNPs were deleterious out of 471 nsSNPs submitted to SIFT in case of DRD3. The 365 nsSNPs of DTNBP1 were checked for tolerance index, showing 20 deleterious nsSNPs. For GRM3 434 nsSNPs were submitted to SIFT for the analysis of tolerance index, resulting in 38 deleterious nsSNPs. Total 1152 nsSNPs for NRG1 were predicted for deleterious nsSNPs, which results in 61 deleterious nsSNPs. PRODH has 426 nsSNPs out of which 31 nsSNPs were deleterious having tolerance index ≤ 0.005. Finally, 286 nsSNPs were searched for deleterious SNPs showing 4 deleterious nsSNPs.

PolyPhen server for damaging SNPs

Polyphon server allows structural level changes in the protein. The nsSNPs of genes were also submitted to the PolyPhen server. PolyPhen searches the damaging SNPs and

uses GRCh37/hg19 as a reference genome. The results by PolyPhen are shown in table 4. PolyPhen uses PSIC (position-specific independent count score) difference of 1.1, nsSNPs above this range is considered deleterious. For protein COMT, two SNPs rs6267 and rs4986871 were found possibly damaging and probably damaging respectively. In the case of DAOA, no SNPs were found damaging and only two SNPs were listed in the PolyPhen as rest of the SNPs were not found in the UniProtKB. In DISC1, five SNPs (rs6675281, rs821616, rs3738400, rs34622148, and rs55795950) were found probably and possibly damaging. DRD3 has only one damaging SNP (rs181422088) according to PolyPhen. For DTNBP1 only one SNP (rs16876589) was found damaging. In the case of GRM3 and NRG1, no SNP and one SNP (rs3924999) were found damaging respectively. For PRODH seven SNPs (rs450046, rs2238731, rs2870984, rs2904551, rs2904552, rs3970559, and rs3970555) were found damaging and in RGS4 there were no damaging SNPs.

Modelling of nine proteins and their damaging mutant structures

The nine genes (*COMT, DISC1, DRD3, RGS4, GRM3, PRODH, NRG1, DTNBP1* and *DAOA*) translate to their respective proteins. Human COMT protein structure was available in protein data bank (4PYI), thus this 3D structure was used as a native structure for COMT mutant structure modelling. The SNPs which were shown deleterious or damaging by both PolyPhen and SIFT server were predicted as functionally important mutations. The two protein mutations occurred at two SNPs rs6267 and rs4986871. The mutations were at position 146 (A→V) and 72 (A→S). These protein mutant structures were modelled using COMT protein as the reference model. COMT protein and its two mutants (A146V and A72S) were uploaded in SwissPDB viewer for the energy minimization. Table 5 shows the total energy after minimization of COMT and its two mutant structures (A146V and A72S) which were found to be -9350.167 kJ/mol, -9302.922 kJ/mol and -9364.405 kJ/mol respectively. The RMSD value of COMT with A146V mutant and A72S mutant was found to be 2.682 A^0 and2.689 A^0 respectively. If the RMSD difference of two protein is higher, the deviation in their structure is also higher thus greater is the change in their functional activity. From table 5, it is clear that the RMSD values of two COMT proteins are higher as compared to native protein, so these SNPs can affect the protein's structure and function. Among two mutants total energy and RMSD values of A72S is greater than A146V mutant. Therefore, A72S mutation is predicted to be more deleterious and affect the functional activity of COMT protein.

The SNPs of DAOA (rs2391191 and rs9558562) were found damaging and deleterious according to Polyphen and SIFT, which may be responsible for altering the function of the protein. The SNP rs2391191 at position 30 where amino acid R (Arginine) substitutes into K (Lysine) and SNP rs9558562 at position 62 where amino acid K (Lysine) substitutes into E (Glutamic Acid). Human DAOA was available in protein data bank (3W4K) but in the bounded form so the protein was modelled using 3W4K model as a reference structure with the help of I-Tasser. The server-generated 2 models and model 2nd was found to be the best model after checking the quality by SAVES 5.0 Server. The predicted model was also checked for Ramachandran Plot which depicts the 92.9% amino acid residues in the most favoured region, 2% in the allowed region and none amino acid was found in the disallowed region. The model had passed all the parameters and hence was used as a native structure for modelling of mutant structures of DAOA protein. The energy of modelled structures was minimized which was found to be -1739.696 kJ/mol for native DAOA, -1800.172 kJ/mol for mutant R30K and -1562.715 kJ/mol for K62E mutant. Also, the RMSD value for R30K with DAOA was 2.43 A^0 and K62E with DAOA was 1.83 A^0. Thus the mutation (R30K) can be predicted to affect the function and structure of the protein as its RMSD value is higher than the native protein (Table 5). In DISC1, the five SNPs (rs6675281, rs821616, rs3738400, rs34622148 and rs55795950) were found deleterious and damaging using Polyphen and SIFT. The SNPs rs6675281, rs821616, rs3738400, rs34622148 and rs55795950 were found at position L607F, S704C, G5V, L330F and T328N respectively. The 3D model of DISC1 was available in protein data bank (5YI4) but in the bounded form so the structure was modelled using 5YI4 as a template. I-Tasser server was used to generate the model and was also checked using the Ramachandran Plot for the stability of the modelled structure. 93.2% residues were found in the favoured region, 3.4% in the allowed region and 3.4% in the outer region thus this model was further used for the analysis. The mutations of DISC1 protein was also modelled using DISC1 as native structure and their energy minimization and RMSD value were calculated. The energy minimization and RMSD value of native DISC1 were found to be -3368.829 kJ/mol and 2.30 A^0 respectively. From the table 5, it is clear that mutants L607F, S704C, T328N have higher RMSD value (2.6 A^0, 2.49 A^0 and 2.52 A^0) from the native DISC1 protein and hence they can affect the protein structure and function. The SNPs of DRD3 were also checked using SIFT and Polyphen and there was only one dbSNP (rs181422088) which was found deleterious and damaging by both the servers. The SNP rs181422088 alters the amino acid valine at position 157 to isoleucine. The 3D model of DRD3 protein was already available in protein data bank (3PBL) so the model was used as a native structure for the mutant modelling. The energy minimization and RMSD value of the mutant structure (V157I) were found higher than the native structure as shown in table 5. The RMSD value and energy minimization of native DRD3 protein were found to be 3.65 A^0 and -18798.566 kJ/mol respectively whereas its mutant form (V157I) have higher energy -19337.301 kJ/

mol and RMSD value 3.89 A^0. Thus the mutant V157I can affect the structure and function of DRD3 protein. DTNBP1 protein has only one SNP (rs16876589) which is found deleterious and damaging by SIFT and Polyphen both. Also, the 3D structure of DTNBP1 was not available in a protein data bank so the structure was modelled using I-Tasser. The structure quality and stability of protein were checked using the Ramachandran Plot. The Ramachandran Plot depicts that the amino acid fall in the favourable region was 94.7%, residues fall in the allowed region was 3.2% and 2.1% in the outer region (Figure 3). The RMSD value and energy after minimization was also predicted which was found to be 2.54 A^0 and -4446.120 kJ/mol respectively for native DNTBP1. These values were found much higher in case of its mutant structure which is 2.78 A^0 and -4570.659 kJ/mol (Table 5). In the case of GRM3 protein, no common SNPs were found to deleterious using both Polyphen and SIFT. Thus, analysis of GRM3 was not performed further. NRG1 protein had one deleterious SNP (rs3924999) common in both Polyphen and SIFT. The rs3924999 substitutes amino acid arginine to glutamine at position 38. The 3D structure of the NRG1 protein was not available in a protein data bank so the protein was modelled using I-Tasser. The structure was further checked for the stability by Ramachandran Plot which plots the amino acid residues, 93.3% of residues were present in the favourable region, 5.7% in the allowed region and 1% in the outer region. The energy minimization and RMSD value of native NRG1 protein were predicted as -4865.479 kJ/mol and 2.3 A^0 respectively. The R38Q mutant of NRG1 protein as energy minimization and RMSD value -5471.062 kJ/mol and 2.9 A^0 which is more than native protein and hence can be predicted to affect the structure and function of the protein. SIFT and Polyphen analysis of PRODH protein results in three common deleterious or damaging SNPs (rs450046, rs2870984 and rs3970555). The rs450046 substitute amino acid glutamine with arginine at position 521, rs2870984 replaces the rionine with methionine at position 446 and rs3970559 substitute amino acid arginine with cysteine at position 453. The 3D structure of PRODH protein was not available in a protein data bank so it was modelled. The structure was further validated for its stability by Ramachandran Plot. The plot predicts that 93.5% of amino acid residues are present in the favourable region and 6.5% are present in the allowed region. Thus this model was used as a native model for the mutant proteins of PRODH. The RMSD and energy minimization of native and mutant proteins was predicted which results in high RMSD values and higher energy minimization (Table 5) than native protein structure. Hence these SNPs can affect the structure and function of the protein. For RGS4 protein, only one deleterious and damaging SNP was found by Polyphen and SIFT. The SNP was rs68678746 which substitute arginine to tryptophan at position 134. The 3D model of RGS4 was already available in protein data bank (1EZT) and the model was used as a native structure against modelling of mutant structure. After the prediction of energy minimization and RMSD values, it was found that SNP rs68678746 can alter the structure and function of protein RGS4. The modelled structure of all the proteins and their respective mutants are shown in figure 2.

Assessment of stabilizing residues among native and mutant structures

S Ride server was used for the identification of stabilizing residues between native and mutant protein structures. The server predicts the stabilizing residue for all nine genes and their respective mutant structures. There was 06 stabilizing residue in native COMT structure as well as in the mutant structure of COMT as highlighted in Table 6. Higher the number of common stabilizing residue the mutation is predicted to be deleterious. Thus both mutations (A146V and A72S) can be an important candidate for schizophrenia caused by COMT protein. In the case of DAOA protein, there was 16 stabilizing residue in native DAOA protein, 14 stabilizing residues in R30K DAOA mutant protein and 13 stabilizing residues in K62E mutant DAOA protein. 14 and 13 stabilizing residues were found common in R30K and K62E mutant DAOA protein respectively. As a higher number of stabilizing residue is common in R30K mutant protein, thus the mutation R30K is predicted to be more deleterious as compared with K62E mutation. Hence the mutation from arginine to lysine at position 30 of DAOA protein is predicted to be more damaging and could be an important candidate for schizophrenia caused by DAOA protein. For DISC1, 04 stabilizing residue was present in the native structure and no stabilizing residues were found in any of the five mutant structures (L607F, S704C, G5V, L330F and T328N). Thus all the five mutant structure could be an important candidate for schizophrenia caused by DISC1 protein. Only one stabilizing residue was found in native DRD3 protein structure and no stabilizing residue was present in its mutant structure (V157I). Hence the mutation can be deleterious and responsible for causing schizophrenia. Two stabilizing residues were found for native DTNBP1 and 05 stabilizing residues were found in G214D mutant structure of DTNBP1, out of which two residues were common. Thus the mutation G214D was predicted to deleterious and can be an important candidate for schizophrenia caused by DTNBP1 protein. In NRG1 protein 06 stabilizing residues were found and no stabilizing residue was found in its mutant structure. 13 stabilizing residues were present in native PRODH protein and 12, 07 and 09 stabilizing residues were found in Q521R, T466M and R453C mutants respectively. Nine, six and nine stabilizing residues were common in Q521R, T466M and R453C mutant structure respectively. As a higher number of stabilizing residue is common in Q521R and R453C mutant protein, thus the mutations Q521R and R453C are predicted to be more deleterious as compared with the T466M mutation.

Hence the mutations are predicted to be more damaging and could be an important candidate for schizophrenia caused by PRODH protein. In native RGS4 and mutant RGS4 protein, there is only one stabilizing residue. Thus R134W mutant protein can be predicted to be deleterious and hence can be an important candidate for schizophrenia caused by RGS4 protein (Table 6).

Screening of antipsychotic drugs

Although there are many FDA approved antipsychotics available but they are only responsible for lowering down the symptoms. The text mining of literature through PubTator suggests that there are seven antipsychotics in priority which are mostly given to patients with schizophrenia.

Docking

The antipsychotics compounds are treated as ligands that were docked against receptors (proteins and their mutant structures). Molecular docking calculates the strength of association and predicts the orientation of molecules when ligands and receptors are bound with each other. The pdb structure of ligands (antipsychotics drugs) and receptor (proteins and their mutated form) were uploaded as input file in PyRx. The studied proteins DTNBP1, COMT, NRG1, PRODH, RGS4, DRD3, DAOA, DISC1 and their damaging mutant forms were docked against drugs aripiprazole, clozapine, fluphenazine, haloperidol, iloperidone, risperidone and lurasidone (Table 7). The native proteins were used as a positive control for drugs against their respective mutant forms. The drug clozapine has a binding affinity of -5.7 kcal/mol with DTNBP1 protein whereas, in case of its mutant form (G214D), glycine at position 214 gets substituted by aspartic acid showed a less binding affinity with clozapine (-6.8 kcal/mol) as compared to the native protein. Thus if protein DTNBP1 gets mutated the antipsychotic drugs binds well with the protein by blocking its active site. So when a diseased person is given antipsychotic drugs it works by lowering the symptoms of the disorder. Likewise, the other drugs in case of mutated protein bind well with each other in comparison with native protein, illustrating that if protein (DTNBP1) mutates (G214D) there is a change in binding energy and the activity of the protein can get blocked, making drug effective. The docked structure of DTNBP1 and DTNBP1_G214 with Lurasidone is shown in figure 4. The protein DRD3 also has one damaging polymorphism i.e, V157I, which means amino acid valine at position 157 gets mutated into isoleucine. The drug which had the least binding activity with DRD3 and its mutant form was risperidone having total energy of -9.4 kcal/mol and -10.4 kcal/mol respectively. The mutant DRD3 protein binds well with the risperidone, blocking the active site of protein thus drugs works well on the protein showing that if protein gets mutated, the drug can bind well with protein. There are two damaging mutants for COMT protein which may affect the pathophysiology of schizophrenia. Risperidone showed the least binding energy with COMT protein (-9.4 kcal/mol) and its mutated form A72S (-10.9 kcal/mol) and A146V (-9.4 kcal/mol). The polymorphism A72S has less binding energy than native protein and polymorphism A146V has similar binding energy as of COMT protein. Thus mutant A72S may be responsible for the cause of schizophrenia as the drugs bind well with the mutant form of COMT. The second mutant A146V does not show much energy deviation illustrating that this mutation may not affect the pathophysiology of the disorder. DAOA is also one of the susceptible genes for schizophrenia which had two deleterious mutations R30K and K62E. In case of native protein DAOA, drug risperidone showed the least binding activity of -8.9 kcal/mol, but its mutant form R30K had less binding energy -9.7 kcal/mol as compared to the native protein. K62E mutant had greater binding energy with clozapine (-8.7 kcal/mol) in comparison with native DAOA protein, but the drug clozapine showed less binding energy (-7.4 kcal/mol) as compared with native protein (-6.2 kcal/mol). Thus both the mutants R30K and K62E may have a role in the cause of schizophrenia as antipsychotic drugs shows better binding with the mutants as compared to the native protein. DISC1 had five damaging mutants (L607F, S704C, G5V, L330F, and T328N) which may have a role in schizophrenia. The drug lurasidone showed better binding energies with mutant form except for T328N as compared with native protein shown in table 7. In the case of T328N drug, haloperidol showed better results in comparison with DISC1 protein. Thus these mutants may also be responsible for schizophrenia. In the case of NRG1 protein, lurasidone shows least binding energy of -6.7 kcal/mol. The mutation R38Q had a damaging effect on the protein also showed least binding energy of -7.0 kcal/mol, which is less than the native protein, thus the polymorphism R38Q may have role in causing the schizophrenia.

PRODH is also one among the genes responsible for schizophrenia which had shown least binding energy with risperidone (-7.5 kcal/mol). There were three deleterious mutations in PRODH (Q521R, T466M and R453C) which showed better binding energies of -8.1 kcal/mol, -9.3 kcal/mol and -10.1kcal/mol respectively against risperidone than PRODH protein. Thus these mutations may have a role in the cause of schizophrenia. In the case of RGS4 only one mutation (R134W) was found to be deleterious, but when antipsychotic drugs were docked against the mutant structure they don't display good results as compared to the native protein. So it can be inferred that the mutant structure R134W may not have many roles in the pathophysiology of the disorder. The docked images of nine proteins and their mutant forms with their respective ligands are shown in Figure 4.

CONCLUSIONS

Schizophrenia is a serious disorder that is gradually posing a threat to human life. Despite enormous research, the cause behind the disorder is unknown. According to the GWAS, there are many genes associated with the disorder. In the current study approaches and servers were used to identify the putative cause behind the disorder. Nine genes were prioritized using the text mining approach. The encoded proteins were checked for disease-associated mutations using SIFT and PolyPhen servers. The proteins and their damaging mutants were modelled and docked with antipsychotics to find the binding energy between them. The drugs were shown to have high binding energy with mutants as compared to native proteins. The interaction energy is considered as best or optimum if the complex is thermodynamically stable with the release of maximum energy thus stabilizing the interaction. Thus from the study, it was concluded that there is a deleterious mutation in the studied nine proteins may be the cause behind the disorder.

ACKNOWLEDGEMENT

The authors express the deep sense of gratitude to the Department of Biotechnology, Graphic Era Deemed to be University for all the support, assistance, and constant encouragements to carry out this work.

Conflict of interest

Authors declare no conflict of interest.

Funding information

No financial support has been received.

REFERENCES

1. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. British journal of pharmacology. 2007;152(1):9-20.
2. WHO (2019). Schizophrenia https://www.who.int/news-room/fact-sheets/detail/schizophrenia. Accessed on 27-08-2020.
3. De Hert M, Correll CU, Bobes J, Cetkovich-Bakmas M, Cohen DA, Asai I, Detraux J, Gautam S, Möller HJ, Ndetei DM, Newcomer JW. Physical illness in patients with severe mental disorders. I. Prevalence, the impact of medications and disparities in health care. World psychiatry. 2011;10(1):52.
4. Gejman PV, Sanders AR, Duan J. The role of genetics in the etiology of schizophrenia. Psychiatric Clinics. 2010;33(1):35-66.
5. Blazer DG, Hernandez LM, editors. Genes, behaviour, and the social environment: Moving beyond nature/nurture debate. National Academies Press; 2006 Dec 7.
6. Vitkup D, Sander C, Church GM. The amino-acid mutational spectrum of human genetic disease. Genome biology. 2003;4(11):1-0.
7. Robert F, Pelletier J. Exploring the impact of single-nucleotide polymorphisms on translation. Frontiers in genetics. 2018 ;30;9:507.
8. Scarff JR, Casey DA. Newer oral atypical antipsychotic agents: a review. Pharmacy and Therapeutics. 2011;36(12):832.
9. Yadav M, Bajaj H, Singh V, & Singh M. Prodrugs: an approach towards better targetting. 2011;3(3).
10. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic acids research. 2001;29(1):308-11.
11. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome research. 2001;11(5):863-74.
12. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic acids research. 2002;30(17):3894-900.
13. Rodrigues JP, Levitt M, Chopra G. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. Nucleic acids research. 2012;40(W1): W323-8.
14. Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. Proceedings of the National Academy of Sciences. 2004;101(18):6957-62.
15. De Lano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsletter on protein crystallography. 2002;40(1):82-92.
16. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. Journal of cheminformatics. 2011;3(1):33.

Table 1: Table showing the list of genes with their accession number

S. No.	Genes	Accession Number	Description
1.	*DTNBP1* (Dystrobrevin Binding Protein 1)	NC_000006.12	Plays a pivotal role in regulating the glutamatergic system. It is located on chromosome 6p22.3.
2.	*DRD3* (Dopamine receptor D3)	NC_000003.12	D3 receptor is mediated by G proteins which inhibit adenylyl cyclase. It is situated at 3q13.31 chromosome.
	COMT (Catechol-O-Methyltransferase)	NC_000022.11	Mammalian enzyme known to be involved in metabolic degradation of catecholamines
	NRG1 (Neuregulin 1)	NC_000008.11	Signalling molecule which has an important role in the organ system growth

Table 1: ***(Continued)***

S. No.	Genes	Accession Number	Description
	GRM3 (Glutamate Metabotropic Receptor 3)	NC_000007.14	Neurotransmitter in the mammalian CNS (Central Nervous system), involved in normal brain functions.
	DAOA (D-amino acid oxidase activator)	NC_000013.11	Functions as an activator of D-amino acid oxidase and are involved in the breakdown of gliotransmitter D-serine
	PRODH (Proline dehydrogenase)	NC_000022.11	Enzyme converting proline into D-1-pyrroline-5-carboxylate
	DISC1 (Distrupted in schizophrenia 1)	NC_000001.11	Gene responsible for mental illness due to its association with dopamine impairments.
	RGS4 (Regulator of G protein signalling 4)	NC_000001.11	Protein which has a role in modulating signalling through G-protein pathways

Table 2: Antipsychotics drugs used against schizophrenia

S. No.	Generic Name	Mode of administration	Recommended dose	FDA status	Indications	PubChem Id
			First-generation antipsychotics			
1.	Haloperidol	Oral, Intramuscular	4-12 mg/d	Approved in 1986	Schizophrenia	CID 3559
2.	Fluphenazine	Oral, Intramuscular	2.5-10 mg/d	Approved in 1960	Schizophrenia and Bipolar disorder	CID 3372
			Second-generation antipsychotics			
3.	Aripiprazole	Oral, Injection	10-15 mg/d	Approved in 2002	Schizophrenia, bipolar disorder	CID 60795
4.	Clozapine	Oral	300-450 mg/d	Approved in 1989	Treatment-resistant schizophrenia	CID 135398737
5.	Iloperidone	Oral	12-24 mg/d	Approved in 2009	Acute schizophrenia	CID 71360
6.	Lurasidone	Oral	40-80 mg/d	Approved in 2010	Schizophrenia	CID 213046
7.	Risperidone	Oral; intramuscular	4-8 mg/d	Approved in 1993	Schizophrenia	CID 5073

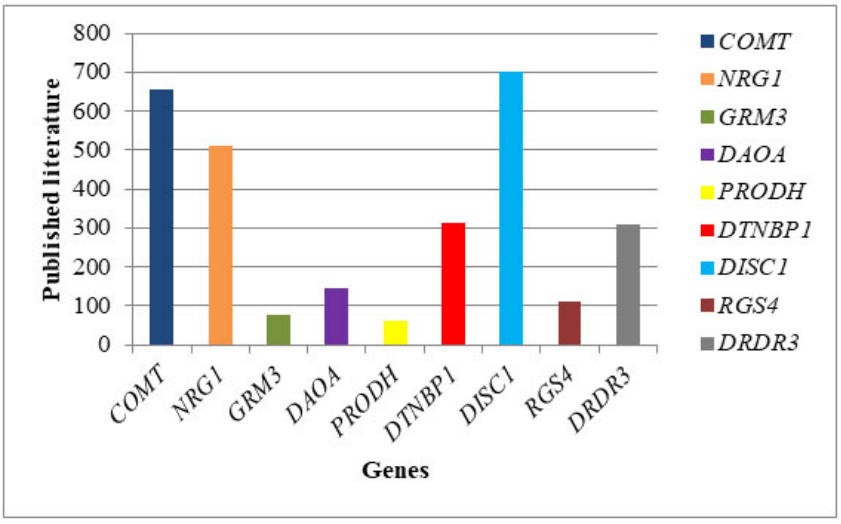

Figure 1: Graph showing all nine genes with their published literature.

Table 3: SIFT result of nsSNPs of the studied genes

dbSNP ID	Nucleotide Change	Amino acid Change	Tolerance index
	COMT		
rs13306281	G/A	V92M	0.002
rs76452330	G/A	D94N	0.005
rs139449932	C/T	R234C	0
rs144463570	C/T	R211W	0.002
rs6267	T/C	A72S	0
rs145561434	C/G	I173M	0.004
rs149909767	G/A	G70R	0.002
rs199710929	C/T	R125C	0.002
rs4986871	A/G	A146V	0
rs200150695	G/A	R184H	0.002
rs201922528	A/T	I104F	0.003
rs373611092	A/G	M90V	0
rs376273380	C/A	A168D	0.001
	DAOA		
rs2391191	G/A	R30K	0
rs367543078	C/G	N42K	0
rs367543079	G/T	V85F	0
rs367543080	C/T	P20S	0
rs367543081	A/G	K74R	0
rs367543081	A/G	K145R	0
rs72549492	C/A	A47D	0
rs72549492	C/A	A118D	0
rs72549492	C/G	A47G	0
rs9558562	C/G	K62E	0
rs72549493	C/G	Q65E	0
rs138223180	G/A	D50N	0
rs138223180	G/A	D121N	0
rs187721661	G/C	R64S	0
rs200207534	T/C	C11R	0
rs200207534	T/C	C82R	0
rs200951630	G/A	G68S	0
rs371012913	C/T	P12S	0
rs371012913	C/T	P83S	0
rs371558248	A/G	D9G	0
rs373343564	C/T	R51C	0
rs373343564	C/T	R122C	0

Table 3: ***(Continued)***

dbSNP ID	Nucleotide Change	Amino acid Change	Tolerance index
	DISC1		
rs66675281	C/T	L607F	0.001
rs367543092	C/T	T573I	0.001
rs367543093	A/G	K577E	0.001
rs28930675	C/T	T453M	0.004
rs34622148	C/T	L330F	0.001
rs76175896	C/T	A83V	0.003
rs76230451	A/G	T561A	0.003
rs78640112	G/T	V350L	0.001
rs34622148	C/T	L330F	0
rs138886515	G/A	E470K	0
rs138886515	G/A	E120K	0
rs139091980	A/G	E161G	0.003
rs139420445	C/T	S216L	0
rs55795950	G/T	T328N	0
rs143165003	C/A	P540Q	0
rs146439119	G/A	R223H	0
rs147158825	C/T	P539L	0
rs148111679	C/T	R569W	0.001
rs821616	G/A	S704C	0.004
rs192018321	C/G	P586A	0.002
rs199530992	C/A	S237Y	0.001
rs199893176	C/G	H256D	0.004
rs200669845	C/G	A530G	0.004
rs201556843	A/C	E236A	0
rs3738400	C/T	G5V	0.001
rs367627719	G/A	G55R	0.001
rs370202687	C/T	T615I	0.003
rs377426796	G/C	A481P	0
	DRD3		
rs76256558	C/G	W85C	0.005
rs141573183	G/A	R148W	0.001
rs143935709	C/T	E57K	0.003
rs144644130	G/A	T14I	0.005
rs148428613	T/C	N342D	0
rs148428613	T/C	N375D	0

Table 3: ***(Continued)***

dbSNP ID	Nucleotide Change	Amino acid Change	Tolerance index
rs149736958	G/A	R254C	0
rs1814220 88	C/T	V157I	0
rs199862630	G/T	P135H	0
rs200010990	G/A	R149C	0
rs200269629	C/G	S117T	0
rs200875766	G/A	R220W	0.004
rs200897022	A/G	I124T	0
rs201102020	C/T	R58Q	0.001
rs201504870	G/T	P178T	0.005
rs201708355	T/A	I364F	0.005
rs201882973	A/G	M52T	0.002
rs201888918	C/G	V334L	0.002
rs202230210	G/A	T155M	0.005
rs368221644	C/T	R149H	0.004
	DTNBP1		
rs370158071	A/G	V207A	0.004
rs77460377	C/G	R54S	0.002
rs367543103	G/C	P317R	0.002
rs367543103	G/C	P318R	0.002
rs367543103	G/C	P161R	0.002
rs16876589	C/T	G214D	0.002
rs16876589	C/T	G215D	0.003
rs142075419	A/G	S230P	0
rs142075419	A/G	S231P	0
rs142075419	A/G	S74P	0
rs144019618	A/G	F255L	0
rs144019618	A/G	F256L	0
rs144019618	A/G	F99L	0
rs147011671	A/G	I218T	0.003
rs147011671	A/G	I219T	0.003
rs200731587	T/A	D174V	0.002
rs200731587	T/A	D175V	0.002
rs200731587	T/A	D18V	0.002
rs201020144	C/A	D329Y	0.001
rs201020144	C/A	D330Y	0.001
rs201020144	C/A	D173Y	0.002

Table 3: ***(Continued)***

dbSNP ID	Nucleotide Change	Amino acid Change	Tolerance index
rs370162147	C/A	G73V	0.004
rs372560190	A/G	L23S	0
rs373060790	C/T	V76M	0.003
rs373182049	T/C	Q98R	0
rs376313138	C/T	V79M	0.001
rs376313138	C/T	V235M	0.002
rs376313138	C/T	V236M	0.002
rs377223155	G/C	S321C	0.004
rs377223155	G/C	S322C	0.004
rs377223155	G/C	S165C	0.005
	GRM3		
rs17856664	C/G	P512A	0.002
rs17856664	C/G	P384A	0.002
rs17856664	C/G	P104A	0.005
rs141671463	C/T	T758M	0
rs141671463	C/T	T630M	0
rs141671463	C/T	T350M	0
rs199660204	T/C	F48S	0
rs199660204	T/C	F456S	0.001
rs199660204	T/C	F328S	0.001
rs200125543	C/T	R66C	0
rs200125543	C/T	R68C	0
rs200125543	C/T	R68C	0.002
rs200125543	C/T	R68C	0.004
rs201158915	C/G	A73G	0
rs201158915	C/G	A75G	0
rs201158915	C/G	A75G	0.005
rs267601607	G/A	E767K	0
rs267601607	G/A	E639K	0
rs267601607	G/A	E359K	0
rs370197727	C/T	R668C	0
rs370197727	C/T	R540C	0
rs370197727	C/T	R260C	0
rs372311811	C/T	R204C	0.001
rs372311811	C/T	R206C	0.001
rs372311811	C/T	R78C	0.003

Table 3: ***(Continued)***

dbSNP ID	Nucleotide Change	Amino acid Change	Tolerance index
rs373913639	G/A	G464R	0.001
rs373913639	G/A	G336R	0.001
rs373913639	G/A	G56R	0.002
rs374144916	G/T	C549F	0.001
rs374144916	G/T	C421F	0.001
rs374144916	G/T	C141F	0.001
rs374569530	C/T	T778M	0.001
rs374569530	C/T	T650M	0.001
rs374569530	C/T	T370M	0.001
rs375977388	T/C	I691T	0.001
rs375977388	T/C	I563T	0.001
rs375977388	T/C	I283T	0.001
rs377189890	G/T	R114S	0.004
	NRG1		
rs367543162	G/T	K116N	0.005
rs367543163	C/G	R250G	0.003
rs367543163	C/G	R229G	0.003
rs367543163	C/G	R279G	0.003
rs367543163	C/G	R352G	0.003
rs367543163	C/G	R287G	0.004
rs367543163	C/G	R122G	0.004
rs367543163	C/G	R276G	0.004
rs3924999	C/T	R38Q	0
rs367543168	C/T	P463L	0
rs367543168	C/T	P586L	0
rs367543168	C/T	P356L	0
rs367543168	C/T	P510L	0.002
rs367543168	C/T	P521L	0.002
rs367543168	C/T	P518L	0.002
rs367543168	C/T	P513L	0.002
rs73672607	C/A	P574H	0.002
rs73672607	C/A	P608H	0.003
rs73672607	C/A	P553H	0.003
rs73672607	C/A	P676H	0.003
rs73672607	C/A	P611H	0.003
rs73672607	C/A	P603H	0.003

Table 3: ***(Continued)***

dbSNP ID	Nucleotide Change	Amino acid Change	Tolerance index
rs76599953	C/T	H306Y	0
rs76599953	C/T	H371Y	0.004
rs76599953	C/T	H141Y	0.004
rs76599953	C/T	H248Y	0.004
rs76599953	C/T	H298Y	0.004
rs76599953	C/T	H295Y	0.005
rs76599953	C/T	H295Y	0.005
rs76599953	C/T	H303Y	0.005
rs76810404	C/A	S136Y	0.005
rs80127039	C/T	R545W	0.001
rs80127039	C/T	R495W	0.001
rs80127039	C/T	R618W	0.001
rs80127039	C/T	R542W	0.001
rs80127039	C/T	R388W	0.001
rs80255389	G/A	V341I	0.002
rs80255389	G/A	V320I	0.003
rs114135581	C/A	N357K	0
rs114135581	C/A	N365K	0
rs114185597	C/T	R619W	0.001
rs114185597	C/T	R389W	0.001
rs114185597	C/T	R543W	0.001
rs114185597	C/T	R554W	0.001
rs114185597	C/T	R551W	0.002
rs114185597	C/T	R546W	0.002
rs115604365	T/G	H24Q	0.002
rs115604365	T/G	H233Q	0.002
rs115604365	T/G	H144Q	0
rs115604365	T/G	H123Q	0
rs115604365	T/G	H246Q	0
rs115604365	T/G	H178Q	0.003
rs116183863	C/A	S523R	0.001
rs139436076	A/C	E422A	0.005
rs139436076	A/G	E422G	0.003
rs141355195	G/A	R619Q	0.005
rs146885321	C/T	R98C	0.002
rs147189312	C/T	R611C	0.003

Table 3: ***(Continued)***

dbSNP ID	Nucleotide Change	Amino acid Change	Tolerance index
rs148350929	G/A	A102T	0.004
rs376791440	C/T	T460M	0.002
rs376169851	G/A	S407N	0.004
	PRODH		
rs142346005	T/C	Y443C	0
rs147233639	G/A	P388L	0.001
rs112389430	G/A	R396C	0.004
rs450046	G/A	Q521R	0.004
rs138400750	C/A	C218F	0.001
rs140831950	C/T	R443Q	0.003
rs140831950	C/T	R335Q	0.003
rs143011525	C/T	R324H	0.002
rs143011525	C/T	R432H	0.003
rs146889635	C/T	V446M	0.001
rs146889635	C/T	V554M	0.001
rs184218784	G/A	R577W	0.004
rs199714362	C/T	D50N	0.005
rs201627713	C/T	A337T	0.002
rs201627713	C/T	A445T	0.003
rs367841908	G/C	H132D	0.005
rs368452830	A/G	L420P	0.003
rs368452830	A/G	L528P	0.003
rs369277468	T/C	N391S	0
rs369277468	T/C	N499S	0
rs370792497	G/A	R323C	0.001
rs370792497	G/A	R431C	0.001
rs2870984	G/A	T466M	0
rs370393004	G/A	R563C	0
rs372030860	A/T	F5Y	0.002
rs372030860	A/T	F113Y	0.002
rs372187772	G/C	Q418E	0
rs372423306	C/T	R399Q	0.003
rs3970559	G/A	R453C	0.005
rs377373292	G/A	R579W	0.005
	RGS4		
rs368678746	A/T	R116W	0

Table 3: *(Continued)*

dbSNP ID	Nucleotide Change	Amino acid Change	Tolerance index
rs368678746	A/T	R134W	0
rs368678746	A/T	R231W	0
rs372256208	G/A	C280Y	0.005

Table 4: List of SNPs for nine genes by using PolyPhen

SNP Id	Position	aa1	aa2	Prediction
		COMT		
rs4680	158	V	M	benign
rs6267	72	A	S	possibly damaging
SNP Id	Position	aa1	aa2	Prediction
rs6270	34	C	S	benign
rs4986871	146	A	V	probably damaging
rs5031015	102	A	T	benign
rs45593642	14	R	L	benign
rs45593642	14	R	L	benign
rs5031015	102	A	T	benign
		DAOA		
rs2391191	30	R	K	possibly damaging
rs9558562	62	K	E	probably damaging
rs3675430 78	42	N	K	benign
rs3675430 79	85	V	F	benign
rs3675430 80	20	P	S	benign
		DISC 1		
rs3738401	264	R	Q	benign
rs6675281	607	L	F	probably damaging
rs821616	704	S	C	probably damaging
rs3738400	5	G	V	possibly damaging
rs34622148	330	L	F	probably damaging
rs55795950	328	T	N	possibly damaging
rs56020408	116	A	V	benign
rs56229136	116	A	V	benign
		DRD3		
rs6280	9	S	G	benign
rs1814220 88	157	V	I	possibly damaging
		DTNBP1		
rs1747045 4	272	P	S	benign
rs1687658 9	214	G	D	probably damaging

Table 4: *(Continued)*

		GRM3		
rs17161026	475	G	D	benign
NRG1				
rs3924999	38	R	Q	possibly damaging
rs1050392	289	M	T	benign
		PRODH		
rs450046	521	Q	R	possibly damaging
rs1807467	455	A	S	benign
rs2238731	427	V	M	possibly damaging
rs2870984	466	T	M	possibly damaging
rs2904551	441	L	P	probably damaging
rs2904552	431	R	H	probably damaging
rs3970559	453	R	C	probably damaging
rs4819756	185	R	W	benign
rs2008720	19	Q	P	benign
rs2870983	472	A	T	benign
rs3970555	406	P	L	probably damaging
		RGS4		
rs14665	195	A	S	Benign
rs368678746	134	R	W	probably damaging

Table 5: Root Mean Square Deviation (RMSD) of native proteins with their respective mutants

	RMSD	Energy after energy minimization
Native COMT protein (4PYI)	2.593 A^o	-9350.167 kJ/mol
A146V with COMT protein	2.682 A^o	-9302.922 kJ/mol
A72S with COMT protein	2.689 A^o	-9364.405 kJ/mol
Native DAOA protein	2.42 A^o	-1739.696 kJ/mol
R30K with DAOA	2.43 A^o	-1800.172 kJ/mol
K62E with DAOA	1.83 A^o	-1562.715 kJ/mol
Native DISC1 protein	2.30 A^o	-3368.829 kJ/mol
L607F with DISC1	2.67 A^o	-3466.654 kJ/mol
S704C with DISC1	2.49 A^o	-3392.598 kJ/mol
G5V with DISC1	1.75 A^o	-1739.696 kJ/mol
L330F with DISC1	1.98 A^o	-3295.540 kJ/mol
T328N with DISC1	2.52 A^o	-3420.069 kJ/mol
Native DRD3 protein	3.65 A^o	-18798.566 kJ/mol
V157I with DRD3	3.89 A^o	-19337.301 kJ/mol
Native DTNBP1 protein	2.54 A^o	-4446.120 kJ/mol
G214D with DTNBP1	2.78 A^o	-4570.659 kJ/mol
Native NRG1 protein	2.3 A^o	-4865.479 kJ/mol

Table 5: *(Continued)*

	RMSD	Energy after energy minimization
R38Q with NRG1	2.9 A^o	-5471.062 kJ/mol
Native PRODH protein	1.52 A^o	-2856.177 kJ/mol
Q521R with PRODH	1.96 A^o	-27330.975 kJ/mol
T466M with PRODH	1.63 A^o	-26642.139 kJ/mol
R453C with PRODH	1.60 A^o	-24870.443 kJ/mol
Native RGS4 protein	1.68 A^o	-6045.847 kJ/mol
R134W with RGS4	1.75 A^o	-6381.265 kJ/mol

* Root Mean Square Deviation

Table 6: Stabilizing residues in native and mutant protein structures

Description	Stabilizing residues
Stabilizing residues in Native COMT protein	Leu112, Leu113, Glu114, Thr138, Glu140, Val188
Stabilizing residues in A146V mutant COMT protein	Leu112, Leu113, Glu114, Thr138, Glu140, Val188
Stabilizing residues in A72S mutant COMT protein	Leu112, Leu113, Glu114, Thr138, Glu140, Val188
Stabilizing residues in Native DAOA protein	Val3, Val4, Val5, Gly7, Lys33, Val34, Ala36, Gly131, Ser136, Leu139, Ile178, Val179, Met203, Phe213, Pro284, Ile306
Stabilizing residues in R30K mutant DAOA protein	Val3, Val4, Val5, Gly7, Lys33, Val34, Ala36, Gly131, Ser136, Leu139, Ile178, Val179, Phe213, Ile306
Stabilizing residues in K62E mutant DAOA protein	Val3, Val4, Val5, Gly7, Lys33, Val34, Ala36, Gly131, Ser136, Met203, Phe213, Pro284, Ile306
Stabilizing residues in Native DISC1 protein	Leu1006, Ser1008, Val1045, Ala1106
Stabilizing residues in L607F mutant DISC1 protein	No stabilizing residue was found
Stabilizing residues in S704C mutant DISC1 protein	No stabilizing residue was found
Stabilizing residues in G5V mutant DISC1 protein	No stabilizing residue was found
Stabilizing residues in L330F mutant DISC1 protein	No stabilizing residue was found
Stabilizing residues in T328N mutant DISC1 protein	No stabilizing residue was found
Stabilizing residues in Native DRD3 protein	Asn47
Stabilizing residues in V157I mutant DRD3 protein	No stabilizing residue was found
Stabilizing residues in Native DTNBP1 protein	Val53, Gly89
Stabilizing residues in G214D mutant DTNBP1 protein	Val53, Gly89, Val156, Val157, Ala159
Stabilizing residues in Native NRG1 protein	Val58, Tyr61, Ile79, Ile80, Gly375, Phe376
Stabilizing residues in R38Q mutant NRG1 protein	No stabilizing residue was found
Stabilizing residues in Native PRODH	Asp61, Ser97, Val130, Arg131, Gly159, Ser164, Arg184, Val222,

Protein	Ala274, Tyr275, Val276, Pro277, Tyr278
Stabilizing residues in Q521R mutant PRODH protein	Asp61, Ile96, Ser97, Val130, Gly159, Ile160, Ser164, Leu183, Arg184, Val222, Ala274, Tyr275
Stabilizing residues in T466M mutant PRODH protein	Asp61, Ile96, Arg131, Gly159, Ser164, Ala274, Tyr275
Stabilizing residues in R453C mutant PRODH protein	Asp61, Ser97, Val130, Arg131, Gly159, Arg184, Val222, Ala274, Tyr275
Stabilizing residues in Native RGS4 protein	Met107
Stabilizing residues in R134W mutant RGS4 protein	Met107

Table 7: Binding energies of studied proteins and their mutant structures against antipsychotic drugs

Protein	Ligand	Binding Affinity (kcal/mol)
DTNBP1	Aripiprazole	-4.9
	Clozapine	-5.7
	Fluphenazine	-5.5
	Haloperidol	-5.1
	Iloperidone	-5.9
	Risperidone	-6
	Lurasidone	**-6.1**
DTNBP1_G214D	Aripiprazole	-5.3
	Clozapine	-6.8
	Fluphenazine	-5.9
	Haloperidol	-5.9
	Iloperidone	-5.4
	Risperidone	-7
	Lurasidone	-6.2
DRD3	Aripiprazole	-7.9
	Clozapine	-8.1
	Fluphenazine	-7.5
	Haloperidol	-9
	Iloperidone	-8.4
	Risperidone	**-9.4**
	Lurasidone	-9.2
DRD3_V157I	Aripiprazole	-7.1
	Clozapine	-8.1

	Fluphenazine	-7.5
	Haloperidol	-8.4
	Iloperidone	-8
	Risperidone	-10.4
	Lurasidone	-8.6
COMT	Aripiprazole	-9.5
	Clozapine	-7.8
	Fluphenazine	-8.4
	Haloperidol	-8.8
	Iloperidone	-8.4
	Risperidone	-9.4
	Lurasidone	-8.9
COMT_A72S	Aripiprazole	-8.1
	Clozapine	-8.8
	Fluphenazine	-7
	Haloperidol	-9.8
	Iloperidone	-9.6
	Risperidone	-10.9
	Lurasidone	-9.9
	Aripiprazole	-9.5
	Clozapine	-7.8
	Fluphenazine	-8.4
	Haloperidol	-8.8
	Iloperidone	-8.4
COMT_A146V	Risperidone	-9.4
	Lurasidone	-8.9
	Aripiprazole	-8.5
	Clozapine	-6.2
	Fluphenazine	-7.9
	Haloperidol	-5.2
	Iloperidone	-6.6
DAOA	Risperidone	-8.9
	Lurasidone	-7.3
	Aripiprazole	-9.3
	Clozapine	-8.1
	Fluphenazine	-6.8
DAOA_R30K	Haloperidol	-5.1
	Iloperidone	-7.6
	Risperidone	-9.7

	Lurasidone	-6.7
DAOA_K62E	Aripiprazole	-8.6
	Clozapine	-7.4
	Fluphenazine	-5.0
	Haloperidol	-6.9
	Iloperidone	-5.2
	Risperidone	-8.7
	Lurasidone	-7.3
DISC1	Aripiprazole	-5.5
	Clozapine	-5.8
	Fluphenazine	-5.8
	Haloperidol	**-6.6**
	Iloperidone	-5.2
	Risperidone	-6.3
	Lurasidone	-6.1
	Aripiprazole	-4.8
	Clozapine	-5.3
	Fluphenazine	-5.3
	Haloperidol	-6.1
	Iloperidone	-5.9
DISC1_L607F	Risperidone	-6.3
	Lurasidone	**-6.4**
	Aripiprazole	-5.6
	Clozapine	-5.7
	Fluphenazine	-5.4
	Haloperidol	-5.6
	Iloperidone	-5.8
DISC1_ S704C	Risperidone	-6.3
	Lurasidone	-6.2
	Aripiprazole	-5.6
	Clozapine	-5.3
	Fluphenazine	-5.7
	Haloperidol	-6
DISC1_ G5V	Iloperidone	-5.9
	Risperidone	-6
	Lurasidone	-6.1
DISC1_ L330F	Aripiprazole	-5.4
	Clozapine	-5.8
	Fluphenazine	-5.5

	Haloperidol	-5.9
	Iloperidone	-6.1
	Risperidone	-6.2
	Lurasidone	-6.4
DISC1_ T328N	Aripiprazole	-5.9
	Clozapine	-5.3
	Fluphenazine	-5.4
	Haloperidol	-6.5
	Iloperidone	-6
	Risperidone	-6.2
	Lurasidone	-6.2
NRG1	Aripiprazole	-6.3
	Clozapine	-6.5
	Fluphenazine	-6
	Haloperidol	-7
	Iloperidone	-6.4
	Risperidone	-7.2
	Lurasidone	-6.7
	Aripiprazole	-6.5
	Clozapine	-5.9
	Fluphenazine	-5.6
	Haloperidol	-6.5
	Iloperidone	-6.5
NRG1_ R38Q	Risperidone	-6.9
	Lurasidone	-7
	Aripiprazole	-7.1
	Clozapine	-6.8
	Fluphenazine	-7.1
	Haloperidol	-7.3
	Iloperidone	-5.8
PRODH	**Risperidone**	-7.5
	Lurasidone	-8
	Aripiprazole	-6.4
	Clozapine	-7.6
	Fluphenazine	-7
	Haloperidol	-7.6
	Iloperidone	-7.4
PRODH_ Q521R	**Risperidone**	-8.1
	Lurasidone	-8.2

PRODH_ T466M	Aripiprazole	-7.1
	Clozapine	-7.3
	Fluphenazine	-7.3
	Haloperidol	-8.3
	Iloperidone	-8.5
	Risperidone	-9.3
	Lurasidone	-9.2
	Aripiprazole	-5.8
	Clozapine	-8.3
	Fluphenazine	-7.3
	Haloperidol	-7.1
	Iloperidone	-6.8
PRODH_ R453C	**Risperidone**	-10.1
	Lurasidone	-7.5
	Aripiprazole	-6.5
	Clozapine	-6.3
	Fluphenazine	-6.4
	Haloperidol	-6.3
RGS4	Iloperidone	-6.7
	Risperidone	-7.3
	Lurasidone	-7.4
	Aripiprazole	-3.7
	Clozapine	-2.2
	Fluphenazine	-4.5
RGS4_R134W	Haloperidol	-2.2
	Iloperidone	-5.2
	Risperidone	-1.7
	Lurasidone	-3.7

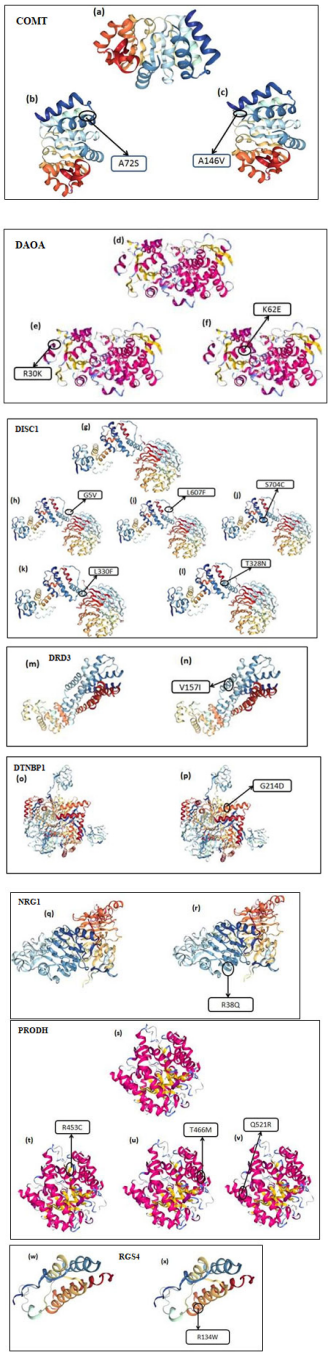

Figure 2: 3D structure of (a) COMT protein (b) Mutated COMT protein A72S (c) Mutated COMT protein A146V (d) DAOA protein (e) Mutated DAOA protein R30K (f) Mutated DAOA protein K62E (g) DISC1 protein (h) Mutated DISC1 protein G5V (i) Mutated DISC1 protein L607F (j) Mutated DISC1 protein S704C (k) Mutated DISC1 protein L330F (l) Mutated DISC1 protein T328N (m) DRD3 protein (n) Mutated DRD3 protein V157I (o) DTNBP1 protein (p) Mutated DTNBP1 protein G214D (q) NRG1 protein (r) Mutated NRG1 protein R38Q (s) PRODH protein (t) Mutated PRODH protein R38Q (u) Mutated PRODH protein T466M (v) Mutated PRODH protein Q521R (w) RGS4 protein (x) Mutated RGS4 proteinR134W

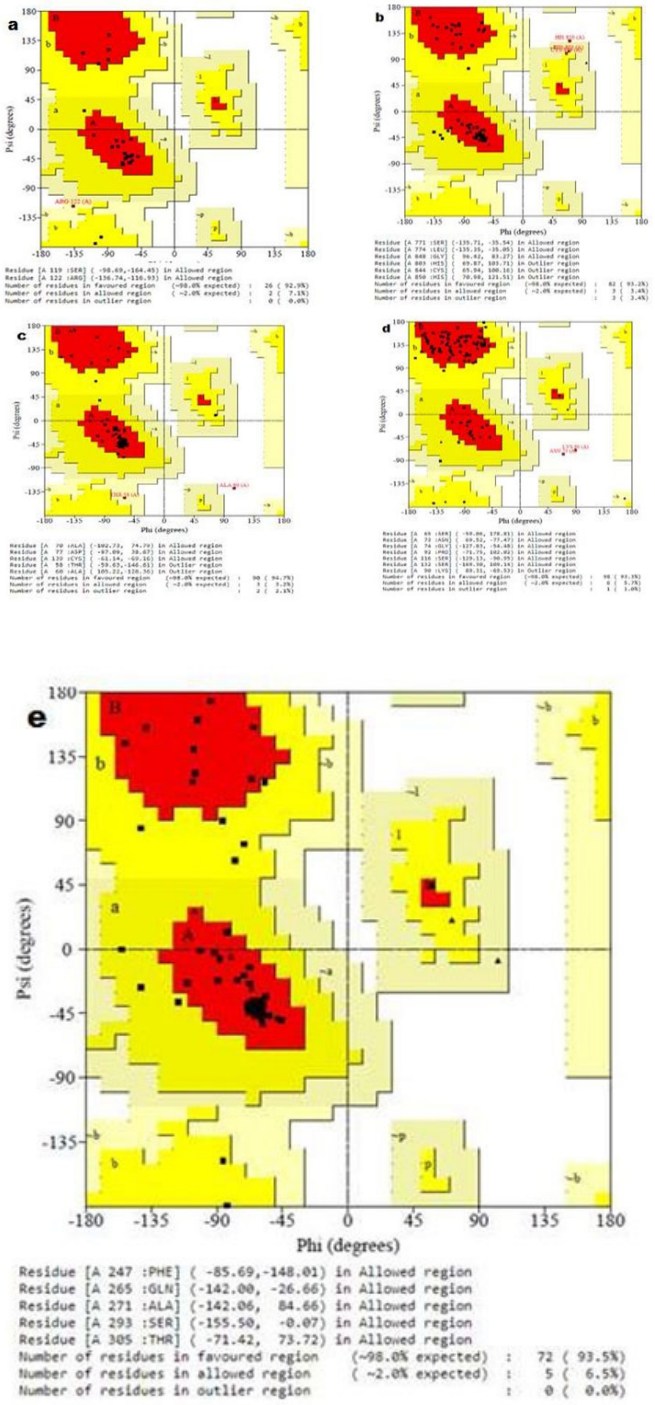

Figure 3: Ramachandran plot of proteins (a) DAOA (b) DISC1 (c) DTNBP1 (d) NRG1 (e) PRODH

Figure 4: Docking images of (a) DRD3 with Risperidone (b) DRD3_V157I with Risperidone (c) DTNBP1 with Lurasidone (d) DTNBP1_G214D with Lurasidone (e) COMT with Risperidone (f) COMT_A146S with Risperidone (g) COMT_A72S with Risperidone (h) DAOA with Risperidone (i) DAOA_K62E with Risperidone (j) DAOA_R30K with Risperidone (k) DISC1 with Lurasidone (l) DISC1_L607F with Lurasidone (m) DISC1_G5V with Lurasidone (n) DISC1_I330F with Lurasidone (o) DISC1_S704C with Lurasidone (p) DISC1_T328N with Lurasidone (q) NRG1 with Lurasidone (r) NRG1_R38Q with Lurasidone (s) PRODH with Resperidone (t) PRODH_Q521R with Risperidone (u) PRODH_R453C with Risperidone (v) PRODH_T466M with Risperidone (w) RGS4 with Lurasidone (x) RGS4_R134W with Lurasidone.