Taylor & Francis
Taylor & Francis Group

Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: www.tandfonline.com/journals/tbsd20

A comprehensive study on identifying the structural and functional SNPs of human neuronal membrane glycoprotein M6A (GPM6A)

Zoya Khalid & Osman Ugur Sezerman

To cite this article: Zoya Khalid & Osman Ugur Sezerman (2021) A comprehensive study on identifying the structural and functional SNPs of human neuronal membrane glycoprotein M6A (GPM6A), Journal of Biomolecular Structure and Dynamics, 39:8, 2693-2701, DOI: 10.1080/07391102.2020.1751712

To link to this article: https://doi.org/10.1080/07391102.2020.1751712

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
2021, VOL. 39, NO. 8, 2693–2701
https://doi.org/10.1080/07391102.2020.1751712

A comprehensive study on identifying the structural and functional SNPs of human neuronal membrane glycoprotein M6A (GPM6A)

Zoya Khalid[a] and Osman Ugur Sezerman[b]

[a]National University of Computers and Emerging Sciences, FAST-NU, Islamabad, Pakistan; [b]Department of Biostatistics and Medical Informatics, Acibadem University, Istanbul, Turkey

Communicated by Ramaswamy H. Sarma

ABSTRACT

Glycoprotein M6A, a stress related gene, plays an important role in synapse and filopodia formation. Filopodia formation is vital for development, immunity, angiogenesis, wound healing and metastasis. In this study, structural and functional analysis of high-risk SNPs associated with Glycoprotein M6-A were evaluated using six different bioinformatics tools. Results classified T210I, T134I, Y153H, I215T, F156L, T160I, I226T, R247W, R178C, W159R, N157S and P151L as deleterious mutants that are crucial for the structure and function of the protein causing malfunction of M6-a and ultimately leads to disease development. The three-dimensional structure of wild-type M6-a and mutant M6-a were also predicted. Furthermore, the effects of high risk substitutions were also analyzed with interaction with valproic acid. Based on structural models obtained, the binding pocket of ligand bound glycoprotein M6-A structure showed few core interacting residues which are different in the mutant models. Among all substitutions, F156L showed complete loss of binding pocket when interacting with valproic acid as compared to the wild type model. Up to the best of our knowledge this is the first comprehensive study where GPM6A mutations were analyzed. The mechanism of action of GPM6A is still not fully defined which limits the understanding of functional details encoding M6-A. Our results may help enlighten some molecular aspects underlying glycoprotein M6-A.

Abbreviations: GPM6A: glycoprotein M6-A; STRING: search tool for the retrieval of interacting genes/proteins; T210I: threonine mutated at position 210 to isoleucine; T134I: threonine mutated at position 134 to isoleucine; Y153H: tyrosine mutated at position 153 to histidine; I215T: isoleucine mutated at position 215 to threonine; F156L: phenylalanine mutated at position 156 to leucine; T160I: threonine mutated at position 160 to isoleucine; I226T: isoleucine mutated at position 226 to threonine; R247W: arginine mutated at position 247 to tryptophan; R178C: arginine mutated at position 247 to cysteine; W159R: tryptophan mutated at position 159 to arginine; N157S: asparagine mutated at position 157 to serine; P151L: proline mutated at position 151 to leucine

ARTICLE HISTORY
Received 20 December 2019
Accepted 31 March 2020

KEYWORDS
GPM6A; glycoprotein; SNPs; non-synonymous; energy minimization; protein stability

1. Introduction

Single nucleotide polymorphisms (SNPs) are the most common form of human genetic variations that are associated with various complex diseases. The largest repository of SNPs is dbSNP (Sherry et al., 2001). Millions of SNPs have been discovered in the coding and non-coding regions of the genome. SNPs can be categorized as synonymous if the variation is not causing the change in amino acid and is non-synonymous if the variation is mutating the amino acid. Overall SNPs occurring more frequently in the non-coding regions than in the coding regions (Khalid & Sezerman, 2014; Mahmud et al., 2016; Spencer et al., 2015). Among all the mutations, non-synonymous substitution occupies 50% of all the SNPs (Gurung & Bhattacharjee, 2018). The deleterious non-synonymous could significantly affect the structure and function of the protein. Such mutations may result in disease phenotypes as well. The non-synonymous SNPs were also studied in various plant varieties that are causing functional changes in the proteins (Khalid et al., 2018).

Neuronal membrane glycoprotein M6-a is a member of tetraspan proteolipid protein (PLP) family and expressed as a stress responsive gene on neuronal membrane proteins in the central nervous system. It was first predicted in mouse brain as an antigen reacting with monoclonal M6 antibodies and causing disruption in the neuronal membranes. In the hippocampus region, the chronic stress downregulates the expression of GPM6A. M6-a plays an important role in synapse and filopodium formation, it has been reported that the knockdown of M6-a decreases the filopodium formation and also affects the synaptic response (Scorticati et al., 2011). Filopodia formation plays an important role in cell migration, immunity, wound healing, adhesion to extracellular matrix and in embryonic development (Garcia et al., 2017; Juliá et al., 2016). Neuronal membrane glycoprotein M6-a binds to

CONTACT Zoya Khalid zoya.khalid@nu.edu.pk National University of Computing and Emerging Sciences, NUCES-FAST, Islamabad, Pakistan.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2020.1751712.

the μ-opioid receptor which also expedites the process of receptor endocytosis which regulates mechanisms like mood, immune response, and stress. Opiate drug is reported as the top drug for treating the chronic stress and pain. The drug receptor binding is regulated by glycoprotein M6-a (Michibata et al., 2009).

The alterations of GPM6A expression level was associated with various complex diseases such as breast cancer (Applewhite et al., 2007), ovarian cancer (Yamazaki et al., 2005) and thyroid cancer (Khalid et al.,2012). Furthermore, the expression level of GPM6A was also associated with other diseases which includes heart failure (Formoso et al., 2015), schizophrenia (Boks et al., 2008), panic disorders (El-Kordi et al., 2013), bipolar disorders (Greenwood et al., 2012), and claustrophobia (El-Kordi et al., 2013). On the other hand, the higher expression of GPM6A has been associated with learning disabilities as reported by Gregor et al. (2014). In addition to that, considering its role in filopodia formation its expression level also disrupted in the brain of depressed suicides as reported (Fuchsova et al., 2015).

Although the mechanism of action of M6-a is poorly comprehended in the research so far still, the significant amount of studies showed its role in filopodia formation. The authors (Rosas et al., 2018) has carried out an in-depth analysis on predicting the significant residues in GPM6A that are involved in filopodia formation. They have tested their method on mammalian cells. The results showed that the c-terminal of GPM6A is highly conserved, hence is involved in filopodia formation. The significance of c-terminal was also been verified with its essential role in the function of other members of tetraspans. Further, few studies have worked on finding the association of non-synonymous SNPs (nsSNPs) with GPM6A. Three variants rs13144140, rs1106568 of GPM6A were reported as associated with schizophrenia (Goes et al., 2015; Soemedi et al., 2012) and rs6846161 with diabetes mellitus (Ripke et al., 2014). No coding amino acids were found for these variants but their association with the diseases were found using GWAS. Also, the authors worked on associating rare copy number variations (CNVs) with M6-a in congenital heart diseases (Soemedi et al., 2012). In addition to that, the authors (Park et al., 2013) had carried out genome wide analysis on diabetes type 2 in Korean population specifically focusing on non-synonymous SNPs in the intronic regions of GPM6A. The de-novo duplication of GPM6A has been reported with various learning behavior abnormalities, also the nsSNPs in the non-coding regions was found associated with various diseases including schizophrenia, claustrophobia and bipolar disorders (Formoso et al.,2015).

Up to the best of our knowledge, no extensive in silico study on GPM6A has been reported before. In view of this consideration, the current study aimed to investigate the impact of deleterious missense mutations on the structure and function of the protein and also to identify the possible drug interactions with M6-a. In silico structural and functional analysis of high-risk SNPs may facilitate the identification of the true genetic markers that are associated with complex diseases.

2. Materials and methods

2.1. Dataset retrieval

The nonsynonymous missense SNPs of neuronal membrane glycoprotein M6-a were collected from NCBI dbSNP (Sherry et al., 2001) https://www.ncbi.nlm.nih.gov/snp/. These SNPs were first analyzed using SNPNexus which has two in-built tools SIFT and PolyPhen that annotates the genomic variants based on seven categories that summarizes conservation, regulatory elements, structural variations, and disease association. SIFT calculates the tolerating index (TI) with value less than 0.05 (TI $<$ 0.05) labelled as intolerant substitutions and tolerant otherwise. Whereas, PolyPhen predictions are based on position specific count score difference which categorizes the mutations as probably damaging, possibly damaging and benign with possibly damaging considered as the most damaging with a calculated score of 1. Next the SNPs were passed to PROVEAN tool that has the cutoff value of –2.5 above which the mutations are considered as neutral while below this threshold value the mutations are labelled as deleterious. Further, the SNPs were subjected to HOPE, meta-SNP, PredictSNP, and SNPs&GO. The SNP is labelled as non-synonymous if it is predicted as deleterious from at least three tools out of six to avoid false predictions. The tools are accessible at http://snp-nexus.org/index.html (Chelala et al.,2009), http://provean.jcvi.org/index.php (Choi & Chan, 2015), http://snps.biofold.org/meta-snp/ (Glaser et al. 2003), https://loschmidt.chemi.muni.cz/predictsnp/ (Calabrese et al., 2009), http://www.cmbi.ru.nl/hope/ (Bendl et al., 2014), http://snps.biofold.org/snps-and-go/snps-and-go.html (Venselaar et al., 2010).

2.2. Prediction of evolutionary conserved residues

The evolutionary conservation of amino acids were analyzed using ConSurf webserver. The server uses Bayesian algorithm to predict the conserved residues along with determining the structural and functional residues. The phylogenetic relations were derived between 50 homologous sequences for determining the degree of conservation among residues. The conservation score of 1–4 is taken as variable, 5–6 is labelled as intermediate, while 7–9 as conserved amino acid position (Glaser et al., 2003).

2.3. Protein contact map prediction

The protein contact map was predicted using NeBcon webserver which is neural network based classifier that uses sequence level information combined with structural information to predict the contact map. The model first calculates posterior probability using Naïve Bayes classifier which then passed to neural networks combined with the features like solvent accessibility, Shannon entropy and secondary structural elements. The server is accessible at https://zhanglab.ccmb.med.umich.edu/NeBcon/ (He et al., 2017).

2.4. Three-dimensional protein structure prediction

The three-dimensional structure of wild-type M6-a and M6-a mutated with the high-risk SNPs were predicted using

Table 1. Deleterious mutations predicted from six different tools.

SNP IDs	Mutation	SNPNexus	PROVEAN	HOPE	Meta-SNP	PredictSNP	SNPs&GO	Result
rs11545193	W141R	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious
rs116256200	T210I	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Deleterious
rs375692732	T134I	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious
rs746939779	Y153H	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Neutral	Deleterious
rs747585562	I215T	Deleterious	Deleterious	Neutral	Deleterious	Deleterious	Neutral	Deleterious
rs750110827	F156L	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Neutral	Deleterious
rs756731427	T160I	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Neutral	Deleterious
rs757535044	I226T	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Deleterious
rs765161647	R247W	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Deleterious
rs768834518	R178C	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious
rs778696357	W159R	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious
rs940555052	N157S	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious
rs959647938	P151L	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious

homology modelling tool I-TASSER. It is a multiple threading approach that accurately predicts structure and function of the protein. The webserver is accessible at https://zhanglab.ccmb.med.umich.edu/I-TASSER/ (Zhang, 2008). The structures were visualized in Chimera 1.11 which is an interactive program for visualization and analysis of molecular data (Pettersen et al., 2004). Further, the model quality of the generated three-dimensional structures were analyzed by ERRAT. Moreover, TM-Align was utilized to compare the wild type and mutant structures. The algorithm generates the RMSD values with higher value indicates more deviation from the wild type structure.

2.5. Predicting protein-ligand binding site

For predicting the binding site of M6-A we have utilized 3DLigandSite server. This server uses ligands of similar structure to predict the binding pocket of the queried protein. Computing conservation score is one of the key feature that has been utilized to compute binding pockets. The server can be accessed via http://www.sbg.bio.ic.ac.uk/3dligandsite/ (Wass et al., 2010).

2.6. Prediction of protein stability

To analyze the stability of GPM6A protein we utilized Mutation Cutoff Scanning Matrix (mCSM) which is a graph based approach. The mCSM works by creating a distance matrix which is computed by finding the pairwise similarity of the atoms present in the vicinity of wild-type residue. The wild-type residue environment is the distance of the atoms from its center point. Further, to this distance matrix the pharmacophore count feature of both wild type and mutant residues are added which is the frequency count of pharmacophore against each 20 residues. This, thus, generated the mCSM signature which is then trained and tested to predict protein stability changes. The tool calculates the change in Gibbs free energy (DDG). We submitted the wild-type structure of GPM6A and the list of mutants to mCSM server for predicting the effect of high-risk SNPs on the protein stability. It is freely accessible at http://biosig.unimelb.edu.au/mcsm/stability (Pires et al., 2014).

2.7. Protein-drug interactions

Drugs interacting with GPM6A were retrieved from CTD and Drug bank. The structures were downloaded from PubChem. BPS-SLIM was utilized to check the drug bound structure of GPM6A with wild and mutant models. This tool is a blind ligand-protein docking method that works best for the low resolution proteins predicted from I-TASSER. The docking scores are calculated based on structural and chemical features similarities between the ligand and the binding pocket. The ligand poses are defined by global similarity of the protein structures and low resolution docking method hence making BPS-SLIM less error prone to the low resolution predicted structures (Lee & Zhang, 2012). The server takes in the structure file of the protein with the ligand structure and predicts five different poses of ligand bound structure of the protein. The BSP-SLIM uses the modified version of SLIM scoring function by adding the conservation of binding site as an additional feature.

2.8. Analyzing protein-protein interactions network

For analyzing the interaction of M6-A with other proteins we have utilized STRING (search tool for the retrieval of interacting genes/proteins) webserver (Szklarczyk et al., 2017). For selecting the interacting partners of M6-A a high confidence threshold of 0.7 was selected. STRING database currently contains 9,643,763 proteins belonging to 2031 different organisms.

3. Results

3.1. Analysis of deleterious substitutions

For identifying the most significant substitutions of GPM6A, SNPs were functionally annotated by utilizing 6 different tools namely SNPNexus, PROVEAN, Meta-SNP, HOPE, SNPs&GO, and Predict-SNP. From dbSNP total 170 unique missense SNPs were retrieved. Further these SNPs were analyzed using SNPNexus which filtered 28 SNPs as non-synonymous substitutions which were further subjected to PROVEAN tool. Among the 28 nonsym SNPs, PROVEAN predicts 13 mutations as the deleterious ones. Further, these shortlisted SNPs were subjected to HOPE, Meta-SNP, PredictSNP, and SNPs&GO. The SNPs that are common in

three tools out of six are selected so that only high-risk SNPs would be analyzed further. The list is tabulated in Table 1.

3.2. Analysis of evolutionary conserved residues

SNPs occurring at conserved regions tend to be more damaging than the SNPs at non-conserved residues. We utilized ConSurf webserver to explore this feature. Residues either buried in the core or exposed at the surface of the protein equally important for analyzing the structure and function of the protein. ConSurf uses evolutionary conserved data and solvent accessibility predictions to determine the structurally and functionally important residues. As there is no experimental or modelled structure available for GPM6A the prediction is based on FASTA sequence. ConSurf generated multiple sequence alignment using MAFFT tool while the homologs were generated using UNIREF90. According to the conservation scores the variants T134I and P151L are highly conserved and buried hence more likely to affect the structure of protein. On the other hand, R178C was found as highly conserved and exposed on the surface of the protein therefore making it functionally important residue. These variants were also predicted as the most deleterious by SNP annotation algorithms hence are classified as the most significant structural and functional residues of GPM6A. The results are tabulated in Table 2 and Figure 1.

Table 2. Conservation profile generated from ConSurf WebServer.

Residue	Conservation score	Prediction
R247W	3	Exposed
W141R	6	Buried
Y153H	3	Buried
T210I	8	Exposed
T134I	9	Highly conserved and buried (s)
I215T	6	Buried
F156L	3	Buried
T160I	2	Buried
I226T	6	Buried
R178C	8	Highly conserved and Exposed (f)
W159R	1	Exposed
N157S	8	Buried
P151L	9	Highly conserved and buried (s)

3.3. Interpretation of protein contact map

The protein contact map of wild-type M6-A and mutant models were predicted from NeBcon webserver. The contact map generates distance between the two residues, the residues are in contact in the sequence if they are in a close proximity in a three-dimensional structure. We have generated separate models of wild type and mutants M6-A to analyze if the mutations are affecting the contact of the residues. The contact maps predicted showed that the mutations are effecting the contacts of M6-A as different residues are in contact when analyzing the mutants as compared to the wild-type M6-A. The details are provided in supporting material.

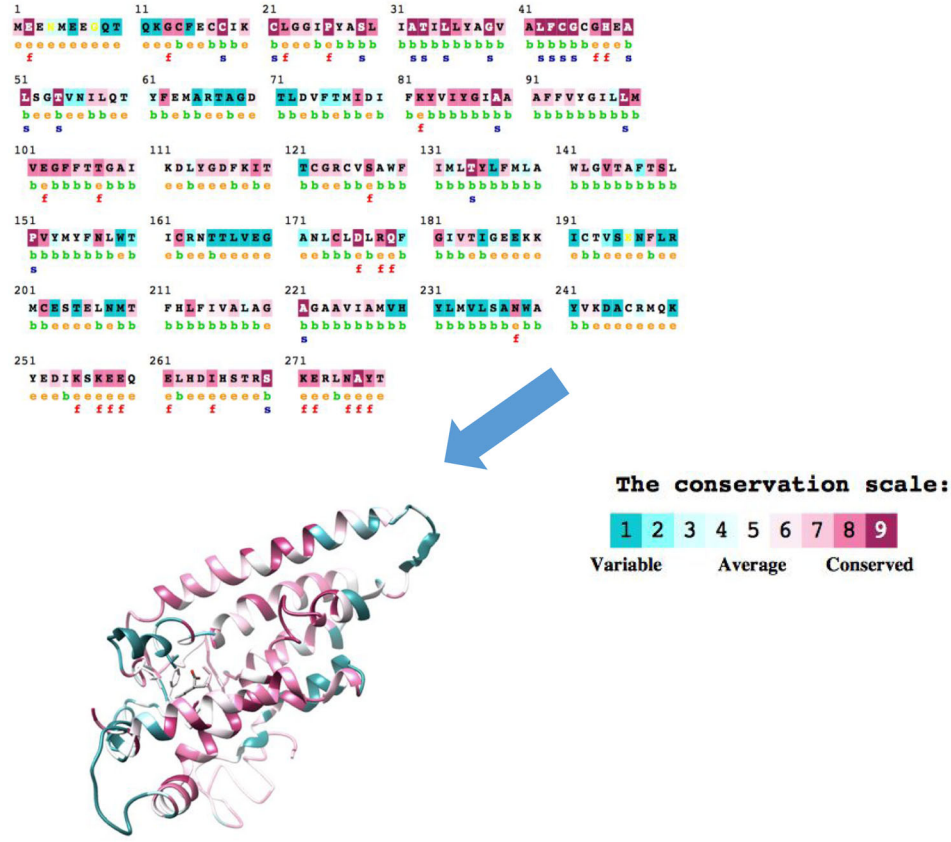

Figure 1. Evolutionary conserved residues predicted from ConSurf sequence and structure mapping. e, An exposed residue according. b, A buried residue. f, A functional residue (highly conserved and exposed). s, A structural residue (highly conserved and buried). X, Insufficient data—the calculation for this site was performed on less than 10% of the sequences.

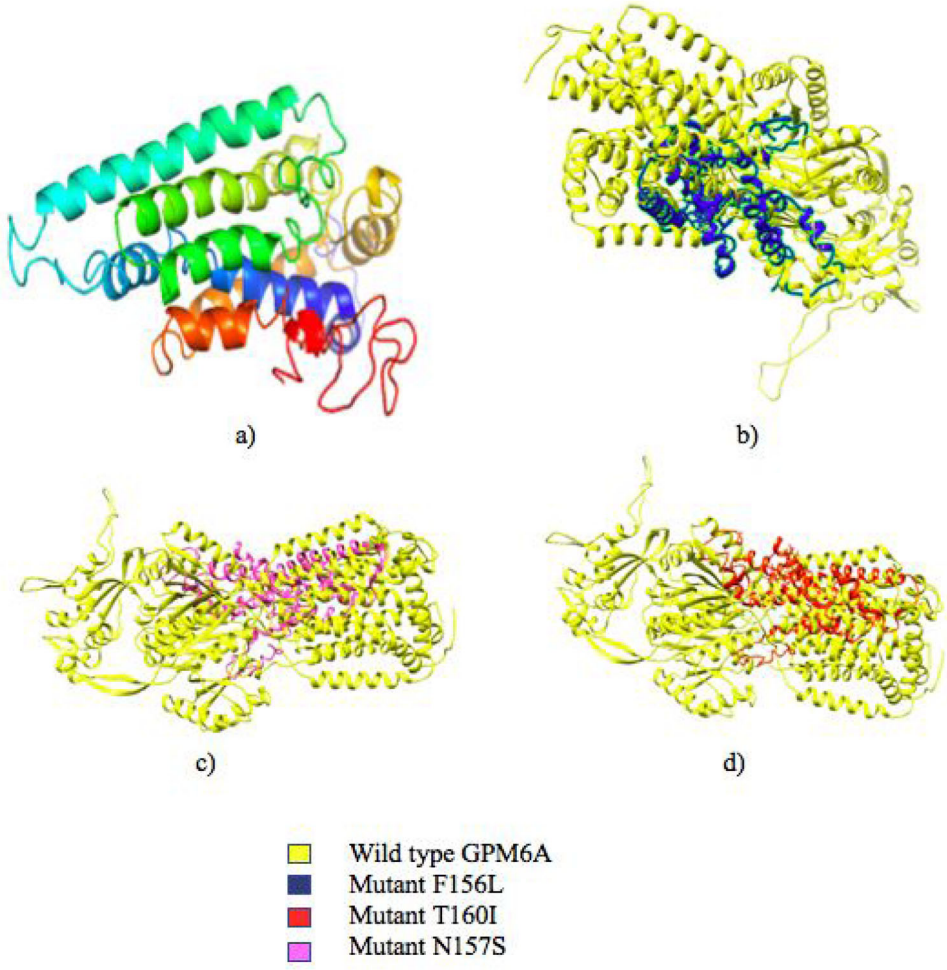

Figure 2. Superimposition of wild type structure with its mutants. (a) The wild-type structure of glycoprotein M6-a generated from I-TASSER. (b) Superimposed model of wild-type M6-a with mutant F156L. (c) Superimposed model of wild-type M6-a with mutant N157S. (d) Superimposed model of wild-type M6-a with mutant T160I.

3.4 Three-dimensional structure prediction with I-TASSER

The tertiary structure of wild type Glycoprotein M6-a protein and the mutant models of M6-a were predicted with homology modelling tool I-TASSER. It uses top 10 threading templates, that covers 73% of the protein, the details of templates are tabulated in supporting material Tables S1 and S2. The generated structures were further viewed by Chimera 1.11 The quality of the three-dimensional model predicted was analyzed by ERRAT server and figured in supporting material Figure S1. The TM-Align determines the RMSD values while comparing the wild type and mutant models. Based on these values we have selected the highly deviated mutant models of F156L, T160I and N157S. These were remodeled using I-TASSER to generate the reliable structures. These mutants models were then superimposed over wild-type protein model. TM-Align generated RMSD value of 5.70 for F156L, 6.36 for T160I and 5.67 for N157S. The superimposed structures are figured in Figure 2.

3.5. Analyzing protein stability upon mutation

For analyzing the effect of deleterious variants on protein stability we utilized mCSM. The mutation is considered as destabilizing if the change in Gibbs energy (DDG) >0.5 and it is considered stabilizing otherwise. The results revealed that all variants decrease the stability of GPM6A protein with F156L, T160I, and N157S as highly destabilizing as predicted. According to some previously reported studies, the decrease in protein stability will likely causes protein misfolding, protein aggregations, degradation and affecting the protein-protein interactions. Hence these polymorphisms are expected to cause maximum damage to the protein structure and function.

3.6. Predicting protein-ligand binding site

The 3DLigandSite server was used to predict the binding site of M6-A protein. It has predicted four different clusters each possess around 10–11 residues. The residues in cluster-1 showed two SNPs at position 156 and 157 of M6-A, which were also predicted as highly destabilizing to the protein function and also interacting with the drug. The results are shown in Figure 3.

3.7. Prediction of protein-drug interactions

The high-risk SNPs were further analyzed for their possible drug-ligand binding activity. More evidences are reported for

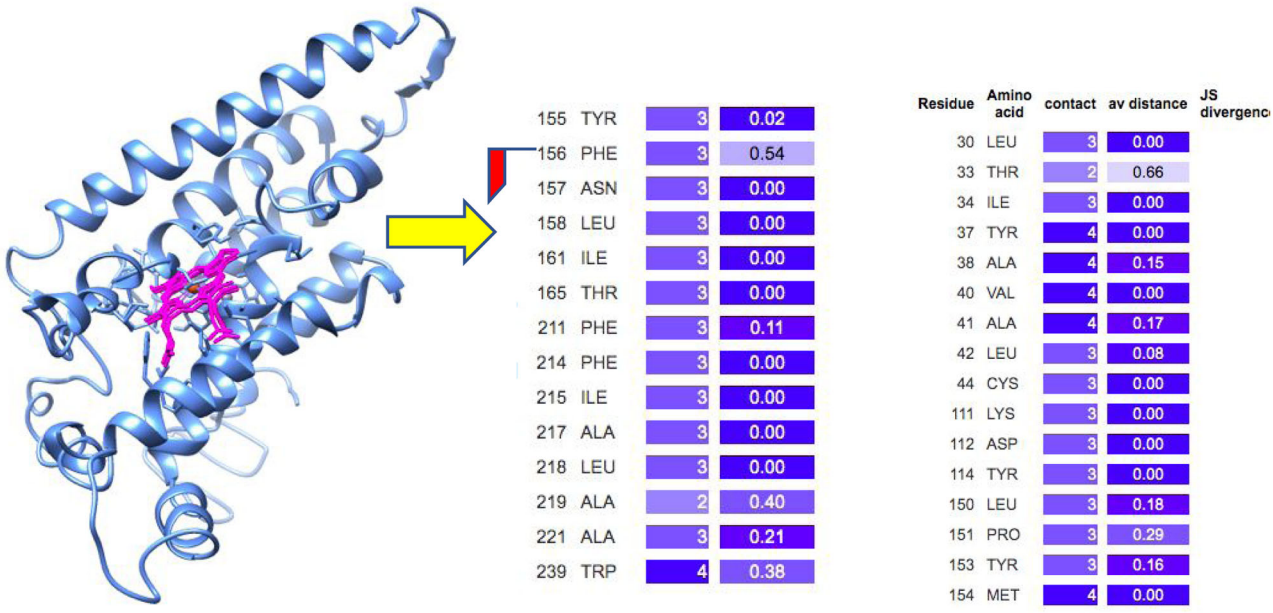

Figure 3. The binding sites predicted from 3DLigandSite. The residues predicted from cluster-1 are shown. Two SNPs are highlighted.

schizopherenia-GPM6A association. No direct interaction of M6-a with any drug found in the literature. Hence, we searched Comparative Toxicogenomics Database (CTD) to look for interactions of M6-a and we obtained the list of drugs interacting with glycoprotein M6-a and picked valproic acid as its showing the highest interaction also this is one of the drugs given to treat schizophrenic patients. Although, the reference does not show any direct evidence of this interaction instead it showed that GPM6A shows increased expression in cells treated with valproic acid. We checked if the SNPs are affecting the binding pocket of the protein by using BSP-SLIM. The ligand structure was obtained from PubChem and was optimized by UCSF chimera for dock preparing the models. Steepest descent algorithm was employed for energy minimization of ligand with wild type structure and three mutant models of M-6A. There were 100 steepest descent steps with 0.02 as a step size. Further, the conjugate gradient steps of 10 with the same step size as 0.02 were selected. These structures were then subjected to BSP-SLIM a total of five poses were generated per model. The final pose selection was obtained by picking the highest score docked model.

All poses were ranked based on the scoring function in order to choose the best scored pose. By analyzing the docking scores obtained we chose the best fitted model with a high score of 4.593 for wild-type M6-A with drug 4.21 of F156L, 4.09 of T160I and 3.978 for N157S. The structures were analyzed with PYMOL and LigPlot. The predicted M6A ligand bound structure showed that Cys162, Leu30, Arg178, Leu175, Phe214, Phe211, Ile215, Ile34 residues interacts with the drug. We checked the binding pocket for the variants and observed that different set of residues interacts with valproic acid. More variability of the residues were found in variants F156L, T160I and N157S. Ligand bounded pocket is predicted by 3D Ligand Site as figured in Figure 3. The variant F156L showed complete loss of the binding pocket hence, this variant is the most crucial one for drug-ligand binding activity (Figure 4).

3.8. Functional characterization using pathway analysis

The STRING database provides the functionally interacting M6-A partners. The predictions are based on experimental data, co-expression and text mining methods. The PPI network generated showed the PPI enrichment value of 7.19e-07 with average clustering coefficient of 0.778. The results are shown in Figure 5. The four KEGG pathways identified by STRING are tabulated in Table 3. The closest interacting partner of M6-A is OPRM1 (Mu-type Opoid receptor) and mutation in it was reported to cause mental disorders specifically schizophrenia. The next best partner is GRIA2 (glutamate ionotropic receptor) which is also reported as associated with depressive and mental disorders exclusively schizophrenia. Hence, loss of protein-protein interactions because of the mutations in M6-A might cause functionality loss and may lead to certain diseases.

4. Discussion

Biological mechanisms underlying complex diseases are quite challenging to understand, as there exists a complex relationship between human genetics and disease traits. Alterations in single residue result in single-nucleotide polymorphisms (SNPs) which are considered as the most significant and common genetic changes. Amongst SNPs, non-synonymous SNPs occurring at coding regions of the genome are highly damaging for the structure and function of the protein which also results in various disease phenotypes. Numerous studies have conducted that predicts the functional consequences of non-synonymous SNPs based on either sequence or structural

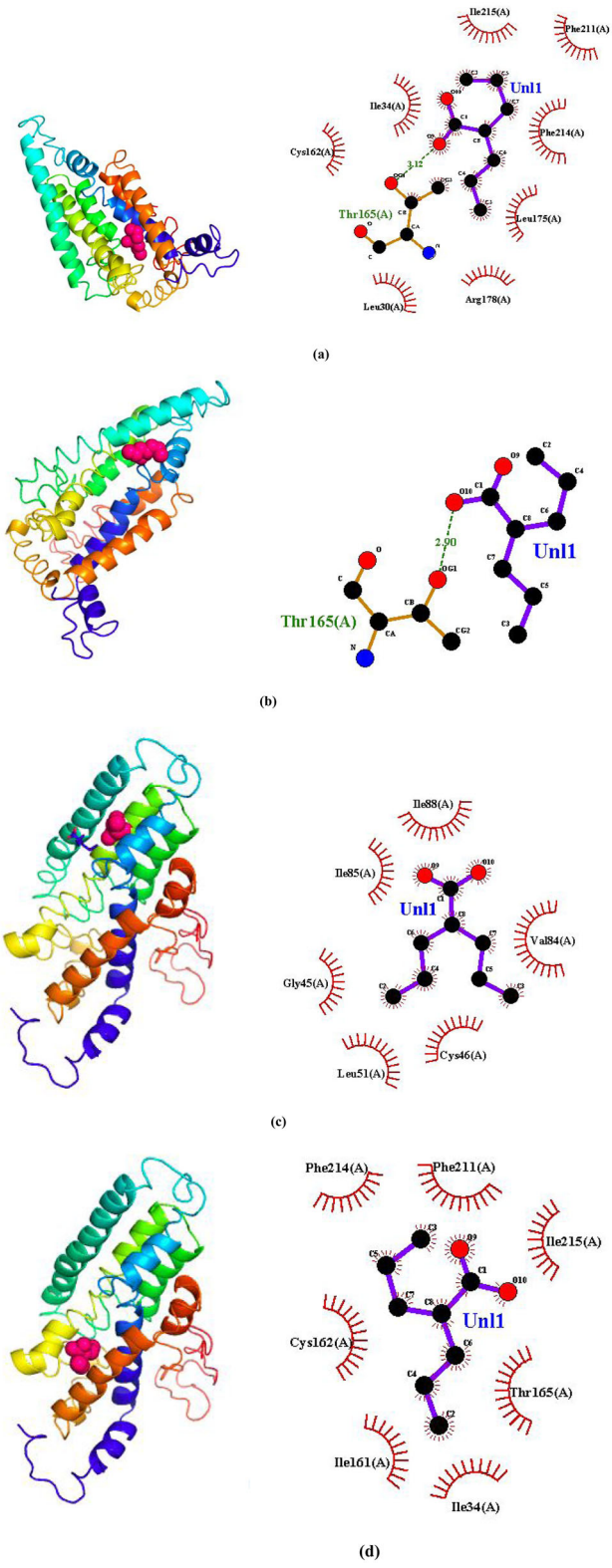

Figure 4. Drug-ligand interactions predicted for GPM6A-valproic acid. (a) The binding pocket of valproic acid when bound to wild type GPM6A. (b) The bound pocket of F156L-Valproic acid which shows complete loss of the binding pocket (c) The bound pocket of T160I-valproic acid that shows different set of residues interacting with the drug (d) The bound pocket of N157S-valproic acid showing different binding pocket for drug valproic acid.

features and in some studies both features were utilized (Arifuzzaman et al., 2020; Arshad et al., 2018).

In the current study, we used various in silico methods to analyze the effect of deleterious mutations on the structure and function of the protein. The 13 most significant mutations W141R, T210I, T134I, Y153H, I215T, F156L, T160I, I226T, R247W, R178C, W159R, N157S, and P151L were identified that were subjected to structure and functional analysis. Among these, the mutants T210I, T134I, R178C, and P151I were predicted as the most conserved residues by ConSurf. Further, the effects of mutants were analyzed on the three-dimensional structure of the protein. The non-synonymous substitutions were either effecting the protein stability or disrupting the conformation of the protein hence will result in poor protein ligand binding. In all the variants identified, the substitutions M108V, F156L, T160I, and N157S were predicted as highly destabilizing by mCSM. Each mutation is superimposed with the wild-type structure to look for possible steric hinderances and clashes due to changes in the side chains.

Further, the structures were also refined by TM-Align, the value of RMSD generated indicates the deviation of mutants from the wild type. Among all, F156L, T160I and N157S were predicted as the most deviated mutants with RMSD of 5.70, 6.36 and 5.67. These variants were further analyzed for their drug ligand binding activity and results showed the loss of binding activity with them. For F156L, drug-ligand bound structure showed complete loss of the binding site as Leucine is more hydrophobic than Phenylalanine and also small in size which might lead to loss of interactions. The mutation introduces a more hydrophobic residue at this position. This can result in loss of hydrogen bonds and/or disturb correct folding which ultimately affects the drug ligand binding activity. For variant T160I, the mutant residue is bigger in size than the wild type also more hydrophobic in nature which will disrupt protein folding. In case of variant N157S, the mutant residue is smaller in size that might leads to the loss of interactions and as its more hydrophobic there are chances that it might leads to protein misfolding.

5. Conclusions

In the current study, the impact of non-synonymous substitutions in GPM6A was predicted by performing bioinformatics analysis. The synonymous substitutions were first separated from non-synonymous ones by using five different tools in order to avoid false predictions. The 13 high-risk SNPs T210I, T134I, Y153H, I215T, F156L, T160I, I226T, R247W, R178C, W159R, N157S and P151L predicted as deleterious were further subjected to structure-function analysis. The three-dimensional structure of M6-A was predicted and analyzed by ERRAT. Taken together, all results of the missense SNPs identified will most likely disrupt the correct folding of the protein which further results in loss of interactions, hence can lead to disease formation. Three mutants F156L, T160I and N157S will most likely disrupt the drug ligand binding activity. The current outcomes can further be experimentally validated in order to determine its clinical significance specifically in the field of pharmacogenomics. Understanding the structural and functional consequences of SNPs would help in early diagnosis and better treatment of complex diseases.

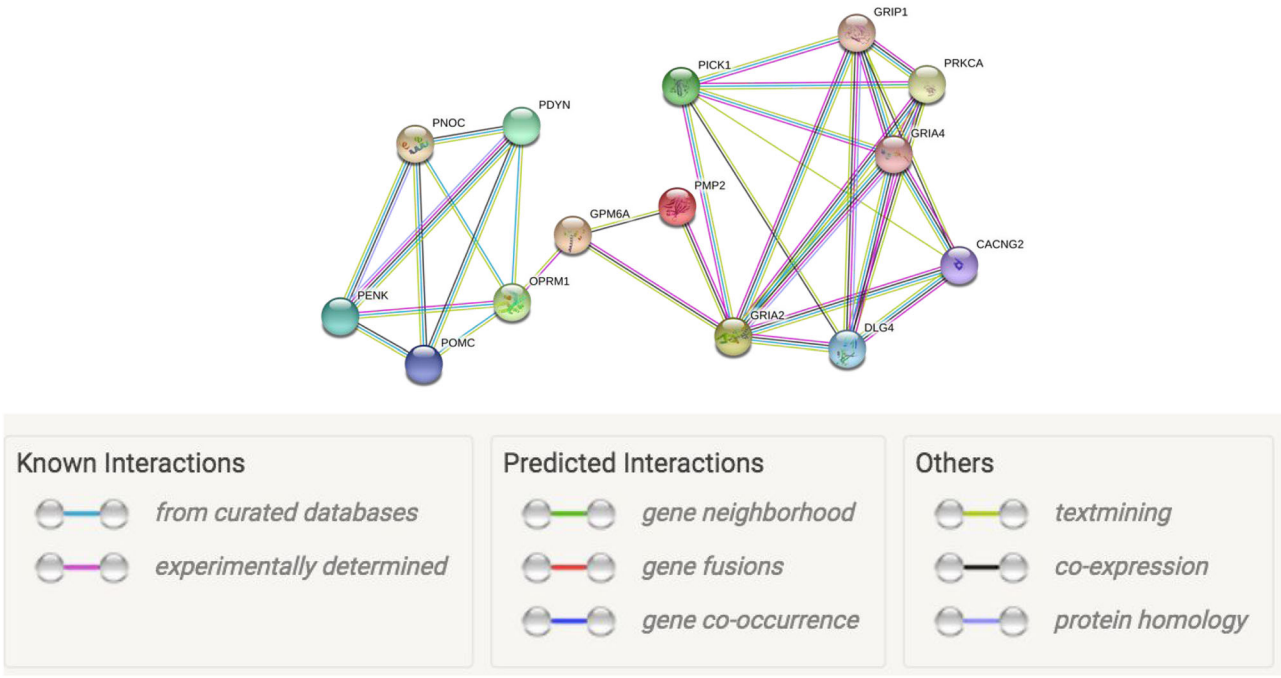

Figure 5. Protein-protein interaction analysis of M6A.

Table 3. Pathway data generated from STRING.

Pathway	Description	Count in genes	FDR
hsa05031	Amphetamine addiction	4 of 65	1.07e-5
hsa04724	Glutamatergic synapse	4 of 112	4.36e-05
hsa05030	Cocaine addiction	3 of 49	0.00016
hsa04713	Circadian entrainment	3 of 93	0.00078

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Disclosure statement

The authors have declared that they have no conflict of interest.

References

Applewhite, D. A., Barzik, M., Kojima, S. I., Svitkina, T. M., Gertler, F. B., & Borisy, G. G. (2007). Ena/VASP proteins have an anti-capping independent function in filopodia formation. *Molecular Biology of the Cell*, *18*(7), 2579–2591. doi:10.1091/mbc.e06-11-0990

Arifuzzaman, M., Mitra, S., Das, R., Hamza, A., Absar, N., & Dash, R. (2020). In silico analysis of nonsynonymous single-nucleotide polymorphisms (nsSNPs) of the SMPX gene. *Annals of Human Genetics*, *84*(1), 54–71. doi:10.1111/ahg.12350

Arshad, M., Bhatti, A., & John, P. (2018). Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. *PLoS One*, *13*(1), e0188143. doi:10.1371/journal.pone.0188143

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., Brezovsky, J., & Damborsky, J. (2014). PredictSNP: Robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Computational Biology*, *10*(1), e1003440. doi:10.1371/journal.pcbi.1003440

Boks, M. P. M., Hoogendoorn, M., Jungerius, B. J., Bakker, S. C., Sommer, I. E., Sinke, R. J., Ophoff, R. A., & Kahn, R. S. (2008). Do mood symptoms subdivide the schizophrenia phenotype? Association of the GMP6A gene with a depression subgroup. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *147*(6), 707–711. doi:10.1002/ajmg.b.30667

Chelala, C., Khan, A., & Lemoine, N. R. (2009). SNPnexus: A web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics*, *25*(5), 655–661.

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), 2745–2747. doi:10.1093/bioinformatics/btv195

Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., & Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human Mutation*, *30*(8), 1237–1244.

El-Kordi, A., Kästner, A., Grube, S., Klugmann, M., Begemann, M., Sperling, S., Hammerschmidt, K., Hammer, C., Stepniak, B., Patzig, J., de Monasterio-Schrader, P., Strenzke, N., Flügge, G., Werner, H. B., Pawlak, R., Nave, K.-A., & Ehrenreich, H. (2013). A single gene defect causing claustrophobia. *Translational Psychiatry*, *3*(4), e254–e254. Aprdoi:10.1038/tp.2013.28

Formoso, K., García, M. D., Frasch, A. C., & Scorticati, C. (2015). Filopodia formation driven by membrane glycoprotein M6a depends on the interaction of its transmembrane domains. *Journal of Neurochemistry*, *134*(3), 499–512.

Fuchsova, B., Julia, A. A., Rizavi, H. S., Frasch, A. C., & Pandey, G. N. (2015). Altered expression of neuroplasticity-related genes in the brain of depressed suicides. *Neuroscience*, *299*, 1–17. doi:10.1016/j.neuroscience.2015.04.057

Garcia, M. D., Formoso, K., Aparicio, G. I., Frasch, A. C., & Scorticati, C. (2017). The membrane glycoprotein M6a endocytic/recycling pathway involves clathrin-mediated endocytosis and affects neuronal synapses. *Frontiers in Molecular Neuroscience*, *10*, 296. doi:10.3389/fnmol.2017.00296

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, *19*(1), 163–164. doi:10.1093/bioinformatics/19.1.163

Goes, F. S., McGrath, J., Avramopoulos, D., Wolyniec, P., Pirooznia, M., Ruczinski, I., Nestadt, G., Kenny, E. E., Vacic, V., Peters, I., Lencz, T., Darvasi, A., Mulle, J. G., Warren, S. T., & Pulver, A. E. (2015). Genome-wide association study of schizophrenia in Ashkenazi Jews. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*(8), 649–659. doi:10.1002/ajmg.b.32349

Greenwood, T. A., Akiskal, H. S., Akiskal, K. K., Study, B. G., & Kelsoe, J. R. (2012). Genome-wide association study of temperament in bipolar disorder reveals significant associations with three novel Loci. *Biological Psychiatry*, *72*(4), 303–310. doi:10.1016/j.biopsych.2012.01.018

Gregor, A., Kramer, J. M., van der Voet, M., Schanze, I., Uebe, S., Donders, R., Reis, A., Schenck, A., & Zweier, C. (2014). Altered GPM6A/M6 dosage impairs cognition and causes phenotypes responsive to cholesterol in human and Drosophila. *Human Mutation*, *35*(12), 1495–1505. doi:10.1002/humu.22697

Gurung, A. B., & Bhattacharjee, A. (2018). Impact of a non-synonymous Q281R polymorphism on structure of human Lipoprotein-Associated Phospholipase A2 (Lp-PLA2). *Journal of Cellular Biochemistry*, *119*(8), 7009–7021. doi:10.1002/jcb.26909

He, B., Mortuza, S. M., Wang, Y., Shen, H. B., & Zhang, Y. (2017). NeBcon: Protein contact map prediction using neural network training coupled with naive Bayes classifiers. *Bioinformatics*, *33*(15), 2296–2306. doi:10.1093/bioinformatics/btx164

Juliá, A. A., Frasch, A. C., & Fuchsova, B. (2016). Neuronal filopodium formation induced by the membrane glycoprotein M6a (Gpm6a) is facilitated by coronin-1a, Rac1, and p21-activated kinase 1 (Pak1). *Journal of Neurochemistry*, *137*(1), 46–61. doi:10.1111/jnc.13552

Khalid, M., Khalid, Z., Gul, A., Amir, R., & Ahmad, Z. (2018). Characterization of wheat cell wall invertase genes associated with drought tolerance in synthetic-derived wheat. *International Journal of Agriculture and Biology*, *20*(12), 2677–2684.

Khalid, Z., & Sezerman, U. (2014). *Interpreting the prevalence of regulatory SNPs in cancers and protein-coding SNPs among non-cancer diseases using GWAS association studies*. International Conference on Applied Informatics for Health and Life Sciences (AIHLS 2014), Kuşadası, Turkey.

Khalid, Z., Sameen, S., Malik, S. I., & Shehzad, S. (2012). Computational analysis on the role of GPM6A in human thyroid cancer. *Journal of Data Mining in Genom Proteomics*, *3*(114), 2153–0602.

Lee, H. S., & Zhang, Y. (2012). BSP-SLIM: A blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins: Structure, Function, and Bioinformatics*, *80*(1), 93–110. doi:10.1002/prot.23165

Mahmud, Z., Malik, S. U. F., Ahmed, J., & Azad, A. K. (2016). Computational analysis of damaging single-nucleotide polymorphisms and their structural and functional impact on the insulin receptor. *BioMed Research International*, *2016*, 1–11. doi:10.1155/2016/2023803

Michibata, H., Okuno, T., Konishi, N., Kyono, K., Wakimoto, K., Aoki, K., Kondo, Y., Takata, K., Kitamura, Y., & Taniguchi, T. (2009). Human GPM6A is associated with differentiation and neuronal migration of neurons derived from human embryonic stem cells. *Stem Cells and Development*, *18*(4), 629–640. doi:10.1089/scd.2008.0215

Park, M.-H., Kwak, S. H., Kim, K. J., Go, M. J., Lee, H.-J., Kim, K.-S., Hwang, J.-Y., Kimm, K., Cho, Y.-M., Kyu Lee, H., Park, K. S., & Lee, J.-Y. (2013). Identification of a genetic locus on chromosome 4q34-35 for type 2 diabetes with overweight. *Experimental & Molecular Medicine*, *45*(2), e7–e7. doi:10.1038/emm.2013.5

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. doi:10.1002/jcc.20084

Pires, D. E., Ascher, D. B., & Blundell, T. L. (2014). mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, *30*(3), 335–342. doi:10.1093/bioinformatics/btt691

Rosas, N. M., Alvarez Juliá, A., Alzuri, S. E., Frasch, A. C., & Fuchsova, B. (2018). Alanine scanning mutagenesis of the C-terminal cytosolic end of Gpm6a identifies key residues essential for the formation of filopodia. *Frontiers in Molecular Neuroscience*, *11*, 314. doi:10.3389/fnmol.2018.00314

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K. H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., Collier, D. A., Huang, H., Pers, T. H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S. A., Begemann, M., Belliveau Jr., R. A. … O'Donovan, M. C.(2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421.

Scorticati, C., Formoso, K., & Frasch, A. C. (2011). Neuronal glycoprotein M6a induces filopodia formation via association with cholesterol-rich lipid rafts. *Journal of Neurochemistry*, *119*(3), 521–531. doi:10.1111/j.1471-4159.2011.07252.x

Soemedi, R., Wilson, I. J., Bentham, J., Darlay, R., Töpf, A., Zelenika, D., Cosgrove, C., Setchfield, K., Thornborough, C., Granados-Riveron, J., Blue, G. M., Breckpot, J., Hellens, S., Zwolinkski, S., Glen, E., Mamasoula, C., Rahman, T. J., Hall, D., Rauch, A., … Keavney, B. D. (2012). Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *The American Journal of Human Genetics*, *91*(3), 489–501. doi:10.1016/j.ajhg.2012.08.003

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1), 308–311. doi:10.1093/nar/29.1.308

Spencer, D. H., Zhang, B., & Pfeifer, J. (2015). Single nucleotide variant detection using next generation sequencing. In *Clinical genomics* (pp. 109–127). Academic Press.

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J., & von Mering, C. (2017). The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*, *45*(D1), D362–D368. doi:10.1093/nar/gkw937

Venselaar, H., Te Beek, T. A., Kuipers, R. K., Hekkelman, M. L., & Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*, *11*(1), 548. doi:10.1186/1471-2105-11-548

Wass, M. N., Kelley, L. A., & Sternberg, M. J. (2010). 3DLigandSite: Predicting ligand-binding sites using similar structures. *Nucleic Acids Research*, *38*(suppl_2), W469–W473. doi:10.1093/nar/gkq406

Yamazaki, D., Kurisu, S., & Takenawa, T. (2005). Regulation of cancer cell motility through actin reorganization. *Cancer Science*, *96*(7), 379–386. doi:10.1111/j.1349-7006.2005.00062.x

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), 40. doi:10.1186/1471-2105-9-40