J. Bangladesh Acad. Sci. 47(1); 53-60: June 2023 DOI:https://doi.org/10.3329/jbas.v47i1.65576

Journal of Bangladesh Academy of Sciences

Journal homepage: http://www.bas.org.bd/publications/jbas.html

Research Article

Computational analysis of functional single nucleotide polymorphisms associated with *NPAS2* gene

Sumaiya Farah Khan*, Fahmida Sultana Rima[1] and Fuad Hasan

Department of Genetic Engineering and Biotechnology, Faculty of Life and Earth Science, Jagannath University, Dhaka, Bangladesh

ARTICLE INFO

Article History

Received: 11 April 2023
Revised: 22 May 2023
Accepted: 11 June 2023

Keywords: Single nucleotide polymorphisms (SNPs), *NPAS2* gene, Circadian rhythm, Computational analysis.

ABSTRACT

Circadian rhythm is a natural physiological process that regulates the sleep-wake cycle and occurs about every 24 hours. The fundamental molecular clock, NPAS2 (Neuronal PAS Domain Protein 2), generates and regulates mammalian circadian rhythms. Various diseases have been linked to single nucleotide polymorphisms (SNPs) in NPAS2. In this study, several computational approaches have been employed to predict the SNPs' functional and structural consequences and investigate the deleterious roles of nsSNPs. Two deleterious nsSNPs (F190L and S105I) were found with predicted functions linking to disease. Alongside, from the obtained ConSurf result, Q73, E61, K51, E167, L317, and R514 were expected as highly conserved and exposed, while T114, T328, and C299 were predicted as highly conserved and buried. However, to establish their role in disease pathogenesis, further studies need to be done on the deleterious mutations of the NPAS2 gene.

Introduction

The largest of the circadian genes and core component of the molecular clock, Neuronal PAS domain protein 2 (NPAS2) is encoded by the NPAS2 gene and resides on chromosome 2 at the band q13. It is a PAS protein 4 (MOP4) member, which acts as a human transcription factor. The amino acid sequence of NPAS2 and Circadian Locomotor Output Cycles Kaput (CLOCK), a transcription factor located in the suprachiasmatic nucleus (SCN), are highly relatable in functional domains (Reick et al., 2001). The NPAS2 gene heterodimerizes with Brain and Muscle ARNT like Protein 1 (BMAL1) and regulates other circadian genes Cryptochrome (Cry1 and Cry2) and Period (Per1, Per2, and Per3)(Albrecht, 2020). The loss of functions of NPAS2 leads to abnormal rhythms. It has several negative impacts, for example, psychiatric diseases like seasonal affective disorder (SAD), bipolar disorder, major depression, and drug addiction, changing patterns of sleep and behavior, and tumorigenesis as well (McClung, 2013; Mukherjee et al., 2010). The immunohistochemistry of the test reveals the presence of NPAS2 protein in both germ cells and Leydig cells. Since Leydig cells are found in the testicle, it functions to produce testosterone in the presence of luteinizing hormone (LH) (Zirkin and Papadopoulos, 2018). In addition, studies have revealed that single nucleotide polymorphisms (SNP) in the NPAS2 gene have been the underlying reason for mood disorders, seasonal affective disorder, and major depressive disorder. However, the mechanisms of NPAS2 in mood-related behaviors are still unclear (Albrecht, 2017 and 2020).

Individual differences are caused by single nucleotide polymorphisms (SNPs), which may be found in gene promoters, exons, and introns, as well as 5'- and 3'-untranslated regions (UTRs). The SNPs occurred in the

*Corresponding author: <sumaiyafarah@yahoo.com>

[1]*Department of Genetic Engineering and Biotechnology, University of Barisal, Bangladesh*

gene coding regions are mainly categorized into synonymous and non-synonymous SNPs (nsSNPs), also called missense SNPs. When mutations do not change the amino acid, it is called a synonymous SNP, whereas non-synonymous SNPs alter the amino acid in humans (Arshad et al., 2018). Substitution of amino acids in protein causes damaging effects, including protein structure destabilization, gene regulation alteration, changes of charge, stability, dynamics, inter/intra protein interactions, which affect the function and structural integrity of the protein (Akter et al., 2022). Following the previous studies, nsSNPs, associated with a large number of genes, have been responsible for about 50% of mutations that are directly implicated in cancer, inflammatory and autoimmune disorders, for instance, Ankylosing Spondylitis (AS), Autoimmune Thyroid Disease/Grave's Disease (AITD), Breast Cancer (BC), Multiple Sclerosis (MS), etc. (Khoruddin et al., 2021).

The risk of human breast and colorectal cancer are associated with a missense SNP in NPAS2 (Ala394Thr). Current research has revealed an association of NPAS2, ARNTL, and CLOCK polymorphisms with a general mood disorder and seasonal affective disorder (SAD) (Kim et al., 2015). In mood disorders, BMAL1 and NPAS2 transcriptionally activate a luciferase receptor in a circadian fashion that directly regulates Maoa transcription (Hampp et al., 2008). NPAS2 is also responsible for the development of anxiety and has a function for the regulation of GABAergic neurotransmission. Additionally, these genes may influence metabolic factors such as body weight and appetite (Albrecht, 2020). So, this study was designed using various bioinformatics tools to discover the most deleterious and damaging nsSNPs of the NPAS2 gene to develop precision medicine for treating the disorders.

Materials and Methods

Retrieval of nsSNPs

All of the nsSNPs in the human NPAS2 gene, as well as their relevant information (reference SNP ID, position, changed amino acid residues and protein accession number, etc.), were collected from NCBI (https://www.ncbi.nlm.nih.gov/). Its protein sequence was also retrieved from UniProtKB (http://www.uniprot.org/uniprot/). Only nsSNPs were chosen for further inspection.

Identification of the deleterious SNPs of NPAS2

To determine the harmful SNPs in the human NPAS2 gene, six different bioinformatics tools (PROVEAN, PolyPhen-2, SNPnexus, PMut, PON- P2, SNAP2) were used. PROVEAN (Protein Variation Effect Analyzer) [http://provean.jcvi.org/ index.php] is a software program that forecasts how an amino acid change or indel would affect a protein's biological function. PolyPhen-2 (Polymorphism Phenotyping v2) [http://genetics.bwh.harvard.edu/pph2/] program makes predictions about the probable effect of an amino acid substitution on the structure and function of a human protein using physical and comparative factors. A web-based tool for variant annotation, SNPnexus [https://www.snp-nexus.org/v4/], makes choosing and prioritizing well-known and new novel genomic variations easier. With high-quality scores, PMut [http://mmb. irbbarcelona.org/PMut] forecasts Mendelian pathogenic mutations. PON-P2 [http://structure.bmc.lu.se/PON-P2/] determines the pathogenicity-association of amino acid substitutions. Using neural networks, SNAP2 [https://github.com/Rostlab/SNAP2] annotates the results of single amino acid alterations in a protein.

Prediction of disease-related SNPs

SNPs & GO [https://snps.biofold.org/snps-and-go/snps-and-go.html] and PhD-SNP [https://snps.biofold.org/phd-snp/phd-snp.htm] were used to predict disease-related SNPs in humans with the NPAS2 gene. With an overall score accuracy of 82%, SNPs&GO (Single nucleotide polymorphism database and gene ontology) is a tool that predicts human disease-related single amino acid variation in protein combined with functional annotations. The predictions of the inputs were divided into neutral and disease in the case of SNPs&GO. With an accuracy of 78% for human proteins, PhD-SNP (Predictor of

human deleterious single nucleotide polymorphisms) is used to predict the disease-associated variants. The PhD-SNP protein prediction result output format is disease and neutral. Together with the reliability index score between 0 and 9, it divides the polymorphisms into disease-associated and neutral categories.

Effect on the human NPAS2 protein stability

By either reducing or enhancing protein stability, SNPs frequently have an impact on protein strength. Several tools were utilized to increase the confidence in the discovered nsSNPs of NPAS2 protein to anticipate these effects. MUpro (http://mupro.proteomics.ics.uci.edu/) and I-Mutant (http://gpcr2.biocomp.unibo.it/I-Mutant.htm) were used to analyze protein stability. During 20-fold cross-validation, MUpro's prediction of changes in protein stability caused by single nucleotide variations in the amino acid sequence has an accuracy of more than 84%. I-Mutant [http://gpcr2.biocomp.unibo.it/I-Mutant.htm], predicts how changes in a single nucleotide will affect a protein's stability. From the protein structure, the I-Mutant predicts protein stability changes upon single-point mutation based on the neural network system.

Phylogenetic conservation of human NPAS2 gene

ConSurf (https://consurf.tau.ac.il/) was used to examine the conservation of the NPAS2 protein sequence. The high throughput characterization of functional protein domains was done using ConSurf. Each amino acid in a protein is assigned a conservation score, ranging from 1 to 9, with scores 1-3 being variable, 4-6 being average conserved, and 7-9 being highly conserved residues. The evolutionary conservation of amino acid positions was determined using protein sequence, and analysis was conducted based on evolutionary relationships between homologous sequences. The highly conserved residues close to high-risk nsSNPs locations were chosen for further study.

Prediction of molecular interaction networks

As proteins are involved in every cellular process, the interactions among them should be maintained to conserve the stability or homeostasis of the system. The online application STRING (https://string-db.org/) was used to forecast the protein-protein interactions. The interacting networks of several proteins with NPAS2 were visualized in PNG format.

Result and Discussion

Retrieval of SNPs

The dbSNP database was used to extract all of the SNPs for the NPAS2 gene. Of the 63,107 SNPs discovered in the human NPAS2 gene, only 663 missense SNPs were chosen for further examination.

Determination of deleterious nsSNPs of NPAS2 gene

Six distinct *in silico* tools were used to examine the detrimental SNPs that could change the function and structure of the NPAS2 protein. This was accomplished using *in silico* tools PROVEAN, SNPNexus, PolyPhen-2, PMut, PON-P2, and SNAP2. Using the PROVEAN software program, 663 missense SNPs were divided into harmful (155 SNPs) and neutral (360 SNPs) [Fig. 1(A)]. SNPNexus projected that 187 nsSNPs were deleterious, 17 were deleterious-low confidence, 265 were tolerated, and 17 were tolerated-low confidence [Figure 1(B)]. The outcome of PolyPhen-2 forecasts nsSNPs that are likely to be damaging, possibly damaging, and benign. This tool identified 154 SNPs as probably damaging, 102 as possibly damaging, and 259 as benign [Figure 1(C)]. The result of PMut divided 146 SNPs as disease and 369 SNPs as neutral [Fig. 1(D)]. The output of PON-P2 revealed 33 SNPs as pathogenic, 201 SNPs as neutral and 281 SNPs' effects were predicted as unknown [Figure 1(E)]. SNAP2 software predicts a score (ranges from -100 strong neutral prediction to +100 strong effect prediction). On this tool, 279 SNPs were predicted to have an impact, and 237 were predicted as neutral [Fig. 1(F)]. In Fig. 2, the percentage of damaging nsSNPs identified by six *in silico* tools are shown.

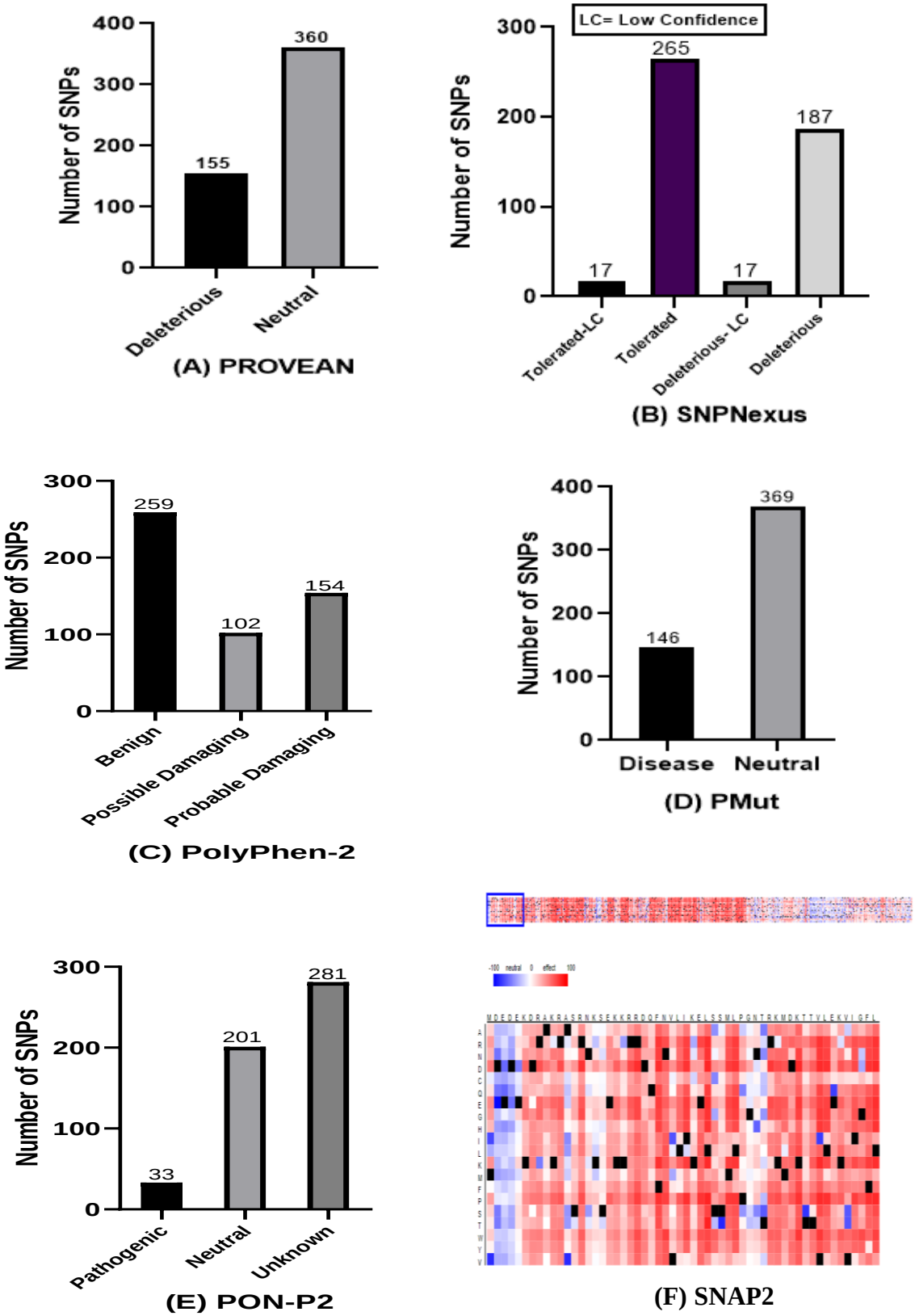

Fig. 1. Graphical representation of the nsSNPs that were explored by the six tools: (A) PROVEAN (B) SNPNexus (C) Polyphen-2 (D) Pmut (E) PON-P2 (F) SNAP2.

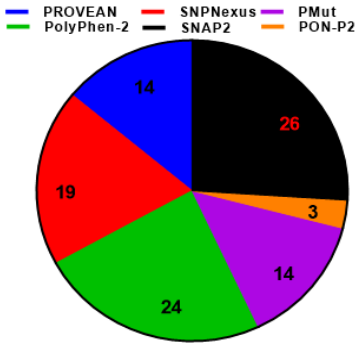

Fig. 2. Pie chart displaying the percentage of damaging nsSNPs identified by six *in silico* tools; PROVEAN, SNPNexus, PolyPhen-2, PON-P2, PMut, and SNAP2.

Analysis of the results from these six tools revealed that 17 SNPs in the NPAS2 protein had been projected to be deleterious (Table 1).

Prediction of human NPAS2 gene disease-related SNPs

Eighty-nine high-confidence nsSNPs were subjected to questions for further analysis to ensure their association with disease or not through SNPs&GO and PhD-SNP servers.

SNPs&GO predicted the impact of the disease-related mutation on the NPAS2 protein function, and it predic-

Table 1. High-risk pathogenic nsSNPs were detected by six *in silico* tools.

rsID	AA change	PROVEAN	SNPNexus (SIFT)	PolyPhen-2	PON-P2	PMut	SNAP2
rs201591122	V810I	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs528544389	Q486H	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs746290665	L220V	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs768357374	M684I	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs775639109	N193K	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs778182952	R570K	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs895884491	Q73K	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs938052839	F190L	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1159238384	E61K	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1161885523	S105I	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1201814308	E158G	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1204687554	A725D	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1247008181	S176G	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1284926566	E774Q	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1305136363	H718Q	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1421402100	K51E	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect
rs1457843152	F666L	Deleterious	Deleterious	Probably damaging	Pathogenic	Disease	effect

ted 14 SNPs as disease and 75 SNPs as neutral [Figure 3(A)]. On the other hand, PhD-SNP predicted 86 SNPs as disease-causing and only 3 SNPs as neutral [Fig. 3(B)]. After analyzing the above result acquired from the six tools and SNPs&GO, PhD-SNP, two SNPs (F190L and S105I) were found to be harmful and disease associated.

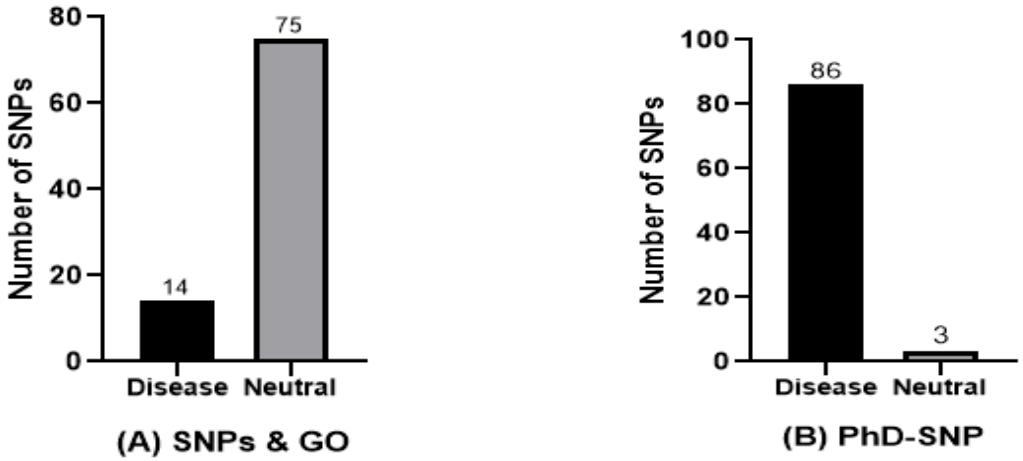

Fig. 3. Graphical representation of disease related SNPs uncovered by (A) SNPs&GO (B) PhD-SNP

Determination of protein structural ability

Due to these SNPs, protein stability may be changed, either lowering or increasing. Increased protein stability causes effective protein action, whereas decreased stability causes hindrance of protein action. A pair of tools I-Mutant and MUpro were used to project these effects, and 106 selected nsSNPs were run through these tools. Since the I-Mutant tool predicts protein structural ability based on the SNP's RI value, it predicted 81 SNPs as decreased stability and 25 as increased stability. According to MUpro, 102 SNPs were predicted to decrease stability and 4 to increase stability.

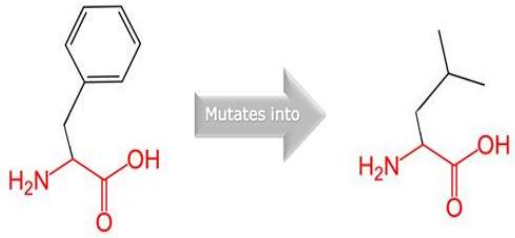

Fig. 4. The schematic structures of the amino acid Phenylalanine (left) and the mutant amino acid, Leucine (right) at 190 positions are shown in the image. Red represents the backbone, which is the same for each amino acid. The unique side chain for each amino acid is colored black.

Alongside, between F190L and S105I, the F190L SNP decreases the stability, and S105I SNP increases the stability of the protein.

Phenylalanine is changed to Leucine at amino acid position 190 due to the F190L SNP mutation (Fig. 4). The mutant residue tends to be in another secondary structure than the wild-type residue, which is anticipated to be in its preferred secondary structure, a β-strand. Therefore the local conformation will be slightly unstable. The altered residue is located in a region crucial for the protein's main activity, so changing the residue might affect the function. The sizes of the wild-type and mutant amino acids also vary, and as the mutant residue is smaller, interactions may be lost as a result.

Analysis of evolutionary conservation of NPAS2

The evolutionary conservation of the 15 most deleterious nsSNPs of NPAS2 protein was calculated by ConSurf. The cutoff value of ConSurf is scaled over nine grades. The higher the conservation score, the higher the chance to be conserved. The ConSurf result of this study predicted that most of the SNPs are highly conserved and exposed, and these highly conserved mutants may have a function in association

with disease. The output of this tool predicted Q73, E61, K51, E167, L317 and R514 as functional residues making them highly conserved and exposed, while T114, T328, and C299 are likely to be structural residues making them highly conserved and buried (Supplementary Fig. 01).

Protein interacting network analysis

The protein interaction network of NPAS2 protein was constructed using the STRING web tool, which also suggested that NPAS2 protein is functionally linked with 10 other proteins, including ARNTL, NR1D1, CRY2, CRY1, BHLHE41, EP300, BHLHE40, RORA, DBP, RXRA (Fig. 5). The entire protein network connection between these 10 proteins might be impacted by changes in NPAS2 gene.

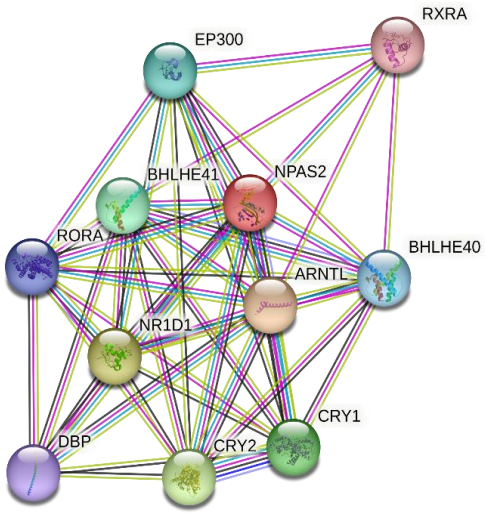

Fig. 5. Protein interaction network of NPAS2 protein

Conclusion

Three known nsSNPs (Ala394Thr, Ser471Leu, and Pro690Ala) in the largest circadian gene, Neuronal PAS domain protein 2 (*NPAS2*), were reported to be involved in breast cancer, indicating a potential array of biomarkers for breast cancer risk (Zhu et al., 2008). Few nsSNPs of the NPAS2 gene are known, while the functional and structural consequences of the nsSNPs in NPAS2 remain unknown. This study identified two potentially deleterious nsSNPs (F190L and S105I) of the NPAS2 gene; further studies are needed to validate this preliminary observation. Molecular dynamics (MD) simulations need to be used to investigate the effects of SNPs on protein structure.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this article.

References

Akter M, Khan SF, Sajib AA and Rima FS. A comprehensive in silico analysis of the deleterious non-synonymous SNPs of human FOXP2 protein. *PLOS ONE*. 2022; 17(8): e0272625.

Albrecht U. Molecular connections between circadian clocks and mood-related behaviors. *J. Mol. Biol.* 2020; 432(12): 3714-3721.

Albrecht U. Molecular Mechanisms in Mood Regulation Involving the Circadian Clock. *Front. Neurol*. 2017; 8: 30.

Arshad M, Bhatti A, and John P. Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. *PLOS ONE*. 2018; 13(1): e0188143.

Hampp G, Ripperger JA, Houben T, Schmutz I, Blex C, Perreau-Lenz S, Brunk I, Spanagel R, Ahnert-Hilger G, Meijer JH and Albrecht U. Regulation of monoamine oxidase A by circadian-clock components implies clock influence on mood. *Curr. Biol.: CB.* 2008; 18(9): 678-683.

Khoruddin NA, Noorizhab MN, Teh LK, Mohd Yusof FZ, and Salleh MZ. Pathogenic nsSNPs that increase the risks of cancers among the Orang Asli and Malays. *Scientific Reports.* 2021; 11(1): 16158.

Kim H-I, Lee H-J, Cho C-H, Kang S-G, Yoon H-K, Park Y-M, Lee S-H, Moon J-H, Song H-M,Lee E, and Kim L. Association of CLOCK, ARNTL, and

NPAS2 gene polymorphisms and seasonal variations in mood and behavior. *Chronobiol. Int.* 2015; 32(6): 785-791.

McClung CA. How might circadian rhythms control mood? Let me count the ways. *Biol. Psychiatry*. 2013; 74(4): 242-249.

Mukherjee S, Coque L, Cao J-L, Kumar J, Chakravarty S, Asaithamby A, Graham A, Gordon E, Enwright JF, DiLeone RJ, Birnbaum SG, Cooper DC, and McClung CA. Knockdown of Clock in the ventral tegmental area through RNA interference results in a mixed state of mania and depression-like behavior. *Biol. Psychiatry*. 2010; 68(6): 503-511.

Reick M, Garcia JA, Dudley C and McKnight SL. NPAS2: An analog of clock operative in the mammalian forebrain. *Science (New York, N.Y.)*. 2001; 293(5529):506-509.

Zhu Y, Stevens RG, Leaderer D, Hoffman A, Holford T, Zhang Y, Brown HN and Zheng T. Non-synonymous polymorphisms in the circadian gene NPAS2 and breast cancer risk. *Breast Cancer Res. Treat.* 2008;107(3): 421-425.

Zirkin BR and Papadopoulos V. Leydig cells: Formation, function, and regulation. *Biol. eprod.* 2018; 99(1): 101-111.