J. Bangladesh Acad. Sci. 49(1); 57-72: June 2025 DOI:https://doi.org/10.3329/jbas.v49i1.71914

Journal of Bangladesh Academy of Sciences

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

Research Article

An in silico approach for the identification of detrimental missense SNPs and their potential impacts on human CRY2 protein

Auroni Semonti Khan, Mahmuda Akter, Mansura Akter Enni and Sumaiya Farah Khan*

Department of Genetic Engineering and Biotechnology, Jagannath University, Bangladesh

ARTICLE INFO

Article History

Received: 18 March 2024
Revised: 17 November 2024
Accepted: 15 January 2025

Keywords: Cryptochrome 2, Circadian clock, FBXL3, Missense SNPs, PER2, MD simulation.

ABSTRACT

Cryptochrome 2 (CRY2) is one of the four proteins of the cell-autonomous molecular clock in mammals. Non-synonymous SNPs of the cry2 gene, resulting in missense variants of CRY2, were correlated with metabolic disorders, cancers, and autism spectrum disorders. This in silico analysis aimed to investigate these missense SNPs having deleterious structural and functional impacts on the human CRY2 protein. Multiple computational tools, homology modeling, and molecular dynamic simulation reported an impact on protein structural stability, function, and binding with PER2 and FBXL3. Our results suggest that missense variants of hCRY2 with L74P, L274P, L309P, F315V, and Y485H mutations were the most destabilizing. These were found to have an overall altered structure, especially in the FAD binding pocket, which in turn can impact the binding of CRY activating compounds, regulatory proteins FBXL3 and PER2. These five missense variants warrant detailed in vitro and in vivo investigations to solidify our findings.

Introduction

The circadian rhythm, or in other words, our internal body clock, keeps our body and its processes with changes in time as well as various environmental stimuli. Cryptochromes (CRYs) are transcriptional repressors constituting the cell-autonomous molecular clock of mammals. These 'peripheral' clocks generate the circadian rhythms in cells through two interlocked transcription-translation feedback loops (TTFL) (Buhr and Takahashi, 2013; Mohawk et al., 2012; Takahashi, 2016), whereas the Suprachiasmatic Nucleus (SCN) in the anterior hypothalamus mediates the circadian rhythm on a macro scale. In the TTFL, the circadian locomotor output cycles kaput (CLOCK), aryl hydrocarbon receptor nuclear translocator-like protein 1 (ARNTL-1) or referred to as BMAL1, form a heterodimer to bind E-box on DNA (CACGTG) causing transcription of PERs and CRYs among other clock-controlled genes (CCGs) in the morning (Gekakis et al., 1998; Hogenesch et al., 1998; King et al., 1997). The proteins PERIODs (PER)s and CRYs then constitute a trimeric complex with casein kinase Iε/δ (CKI) in the cytoplasm to translocate into the nucleus in the late afternoon or evening. The transcription of CCGs is inhibited by the interaction of this trimeric complex and BMAL1/CLOCK (Vielhaber et al., 2001). Polyubiquitination and subsequent CRY1, CRY2, PER1, and PER2 degradation occur by specific E3 ligase complexes (FBXL3 and β-TrCP). Eventually, their negative feedback repression is lifted as these proteins decrease, and CLOCK–BMAL1 transcription activity begins again, completing one cycle of the TTFL (Buhr and Takahashi, 2013).

CRYs (CRY1 and CRY2) are flavoproteins (Partch and Sancar, 2005; Reppert and Weaver, 2002). Mammalian cryptochromes, weighing 66.9kDa, are composed of a highly conserved photolyase-homology

*Corresponding author:<sumaiyafarah@yahoo.com>

region (PHR) and a unique Cryptochrome C-terminal (CCT) extension (CCE). This protein of 593 amino acids has a FAD-binding pocket in this PHR (Xing et al., 2013). Another secondary pocket is present, which helps CRYs bind to CLOCK: BMAL1 complex (Fribourgh et al., 2020; Rosensweig et al., 2018). CRY1 and CRY2 have similar functions but distinct roles, especially in determining the rhythm of the SCN biological clock and its heterogeneous circadian transcriptional outputs. Various enzymes regulate CRY2 activity, and it also undergoes PTM, including phosphorylation and ubiquitylation, which affects its stability (Reischl and Kramer, 2011; Stojkovic et al., 2014). CRY2 has been demonstrated to cause the degradation of c-MYC (Huber et al., 2016), TLK2 (Correia et al., 2019), E2F family members (Chan et al., 2020) by its interaction with SCFFBXL3 (SKP1-CUL1-F-box protein complex having the F-box protein FBXL3) which ubiquitinates these proteins. On the contrary, PER proteins protect CRY from degradation by FBXL3 (Xing et al., 2013). More importantly, PER2 is one of the most important binding partners of CRY2. Their binding mediates the TTFL in cellular circadian clocks, completing the negative phases of the loop. Interruption of circadian oscillations increases with the association of both Cry1 and Cry2 genes with several types of cancer. Genome sequences from the Cancer Genome Atlas for six global and eight subcontinental ancestries report that among other CCGs, CRY2 is resistant to mutation in healthy somatic tissues making its cancer-associated mutations functionally important (Chan et al., 2020). Multiple studies have revealed the involvement of single nucleotide polymorphisms (SNPs) in the human Cry2 gene (hCry2) with prostate and breast cancer (Zienolddiny et al., 2013) and Non-Hodgkin's lymphoma (Hoffman et al., 2009). Cry2 gene polymorphism has also been shown to be associated with bipolar disorder, depression (Lavebratt et al., 2010a), attention-deficit/hyperactivity disorder (ADHD) (Wang et al., 2020), and dysthymia (Kovanen et al., 2014). SNPs in hCry2 have also been associated with metabolic disorders such as diabetes, dyslipidemia (Dupuis et al., 2010; Kelly et al., 2012; Salazar et al., 2021), cardiovascular disorders (Škrlec et al., 2018), and the more obvious, sleep disorder (Hirano et al., 2016) as well as osteoporosis (Li et al., 2016), affective disorders (Kripke et al., 2009).

SNPs are single-base nucleotide substitutions that are prevalent in less than 99% of a population (Schork et al., 2000). Among various types, nonsynonymous SNPs (nsSNPs) result in missense mutations in the coding region, causing aberrations of protein structure, stability, and function as well as binding to respective partners in cellular context (Chasman and Adams, 2001; Kucukkal et al., 2015; Petukh et al., 2015). These nsSNPs and resulting missense proteins have been linked to hereditary diseases, metabolic disorders, and diseases related to circadian rhythms, as stated above. However, isolating the nsSNPs that contribute to diseases is highly challenging since it requires numerous experiments of thousands of SNPs in a potential candidate gene. In contrast, the use of bioinformatics to comprehend the potential impact of nsSNPs has been reported (Kumar and Purohit, 2012; Rajendran et al., 2018; Rajendran and Sethumadhavan, 2014). This study, therefore, aims to detect detrimental non-synonymous SNPs that have pathogenic effects on the hCRY2 protein among the copious number of missense variants. Several computational tools were utilized to investigate the impact of the nsSNPs on protein structure as well as its binding to other protein partners- FBXL3 and PER2. Apart from these calculative, descriptive, and inferential analyses, molecular dynamics (MD) simulation was applied to get insights into the impact of damaging variations in the native tertiary structure of the candidate protein.

Materials and Methods

Retrieving SNPs of human CRY2 protein (hCRY2)

Human CRY2 protein was looked up in dbSNP (Sherry et al., 2001). About 13088 SNPs information were available in the database, among which the 436 nonsynonymous SNPs were chosen and retrieved for this study. The amino acid sequence of hCRY2 was collected from the UniProt Knowledge Base (UniProtKB) (Accession Number: NP_066940.3). The following sections depict the series of analyses to conclude the missense variants of hCRY2, which are summarized in a flowchart (Fig.1).

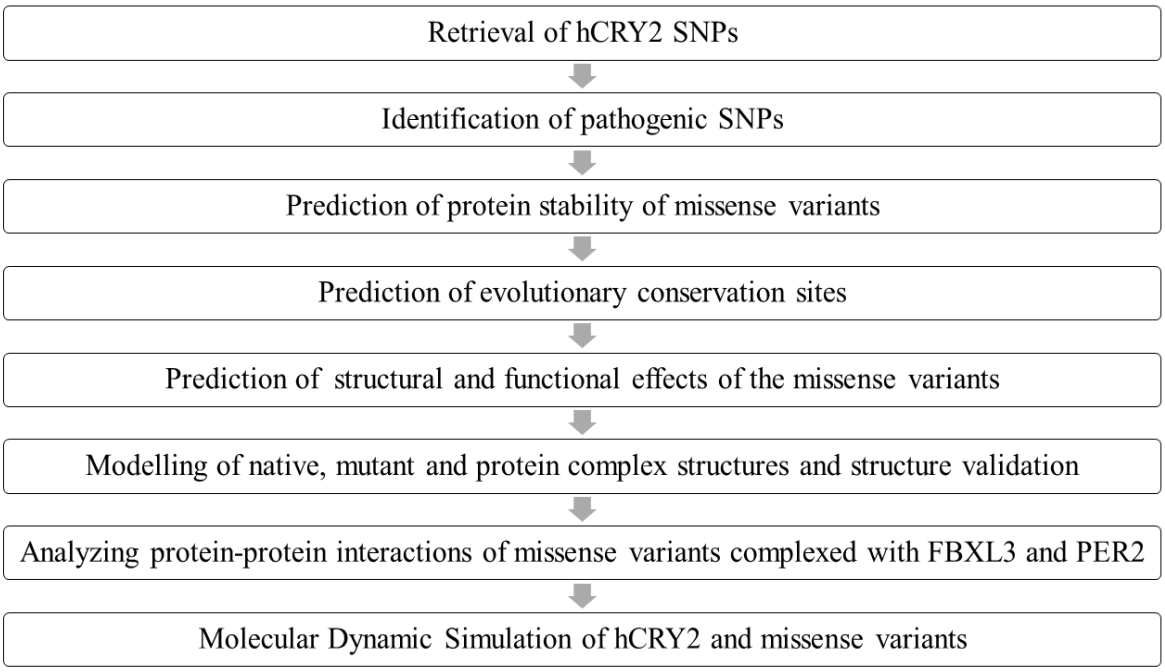

Fig. 1. Flowchart of the study design.

Identifying the most detrimental SNPs

Deleterious, pathogenic and disease-causing SNPs were predicted using Polyphen-2 (Adzhubei et al., 2013), SNAP2 (Hecht et al., 2015), PON-P2 (Niroula et al., 2015), Rhapsody (Ponzoni et al., 2020), PMut (López-Ferrando et al., 2017) and SNPs&GO (Calabrese et al., 2009). PolyPhen-2 (Polymorphism Phenotyping version 2) investigates the damaging missense variations to assess the structural and functional impacts of single amino acid substitution on human protein (Adzhubei et al., 2013). SNAP2 (Screening for non-acceptable polymorphisms) anticipates the functional effects of an amino acid alteration by using a neural network (Hecht et al., 2015). To assess the pathogenicity of the variants, PON-P2 was utilized to assort the variants into unknown, neutral, and pathogenic classes based on a random forest probability score (Niroula et al., 2015). Using machine learning, Rhapsody predicts the pathogenicity based on protein sequence, structure, and dynamics and classifies them into deleterious or neutral (Ponzoni et al., 2020). To address the association of SNPs with disease or not, PMut (Pathogenic Mutation prediction) exploits neural networks to predict evolutionary and structural properties (López-Ferrando et al., 2017). SNPs&GO is based on the support vector machines (SVM) method that predicts whether a given mutation can cause a disease or not, utilizing gene ontology (GO) annotation (Calabrese et al., 2009).

Predicting change in protein stability due to missense variants

MUpro (Cheng et al., 2006), mCSM (Pires et al., 2014), DeepDDG (Cao et al., 2019), INPS3D (Savojardo et al., 2016) and PremPS (Chan et al., 2020) were used to determine the effects of the variants on the stability of hCRY2 protein structure. MUpro utilizes sequence and structure information to assess the effect of point mutations (Cheng et al., 2006). mCSM considers the distance profiles between atoms to predict whether mutations destabilize the protein (Pires et al., 2014). Online tools, DeepDDG, INPS3D, and PremPS, were applied to estimate the impact of thermodynamic changes ($\Delta\Delta G$) on the stability of the protein induced by a missense mutation.

***Analyzing evolutionary conservation sites in* hCRY2**

The ConSurf tool was used to analyze the evolutionary pattern of the protein's amino acids to find functional regions by studying the evolutionary dynamics of amino acid substitution. The conservation score is calculated from the Bayesian calculation method on a scale of 1 to 9; each amino acid is scored as variable, intermediate, and conserved, showing its degree of conservation (Ashkenazy et al., 2016).

Predicting structural and functional effects of the missense variants

The structural analyses web server Project HOPE (http://www.cmbi.ru.nl/hope/home) was used for this. Amino acid sequences of the native hCRY2 and those of the mutant variants were input in this server. The three-dimensional tertiary structures of the proteins deposited in the Distributed Annotation system (DAS) servers and Uniprot database are used by Project HOPE for evaluating the structural impacts of substitution mutations on the native protein (Venselaar et al., 2010).

Determination of native and mutant structures of* hCRY2 *and structure validation

The native and mutant structures of the hCRY2 protein were generated by SWISS-MODEL (Waterhouse et al., 2018) through homology modeling using murine CRY2 (PDB ID: 4i6j.1.A) as a template. SWISS-MODEL is a web-based server that relies on template-based modeling to generate protein structures as well as homomeric and heteromeric complexes. The structures were validated using the ERRAT (Colovos and Yeates, 1993) and PROCHECK (Laskowski et al., 1996) tools in the SAVES 6.0 server.

Analyzing impacts on protein-protein interactions in missense variants

The crystal structure of murine CRY2 in its FBXL3–SKP1-complexed form (PDB ID: 4i6j.1) was found in existing literature (Xing et al., 2013). Based on this template, homology modeling was performed to form a heterodimer complex of hCRY2 and FBXL3 using SWISS MODEL (Waterhouse et al., 2018). A similar method was followed to generate the hCRY2 and PER2 complex, using a published (Nangle et al., 2014) crystal structure of murine CRY2 and PER2 in complex (PDB ID: 4u8h.1). Then, the interactions of hCRY2 with its binding partners FBXL3 and PER2 were analyzed to observe any changes in protein interactions due to missense mutations. The computational tools mCSM-PPI2 (Rodrigues et al., 2019), MutaBind2 (Zhang et al., 2020), SAAMBE-3D (Pahari et al., 2020) and BeAtMuSiC V1.0 (Dehouck et al., 2013) were applied to explore any variation in protein interactions. MCSM-PPI2 utilizes optimized graph-based signatures to model the effects of point mutations in protein-protein binding affinity, focusing on the inter-residue non-covalent interaction network (Rodrigues et al., 2019). MutaBind2 predicts changes in binding affinity between proteins in missense variants through statistical potentials, molecular mechanics force fields, and fast side-chain optimization algorithms made via random forest (RF) formula to assess alterations (Zhang et al., 2020). SAAMBE-3D (Single Amino Acid Mutation Change of Binding Energy) exploits machine learning to analyze alterations in protein-protein interactions as an effect of single amino acid mutation (Pahari et al., 2020). BeAtMuSiC V1.0 calculates the changes in binding free energy of the protein-protein interactions through the combination of the effect of the mutation on the strength of the interactions and the gross stability of the complex (Dehouck et al., 2013).

Simulating the dynamics of* hCRY2 *and its selected variants

CABS-flex 2.0, a coarse-grained protein model, was used to study the structural flexibility of the wild-type and mutant 3D models of hCRY2 (Kuriata et al., 2018). Three-dimensional models of the hCRY2 missense variants were produced using template-based modeling in SWISS MODEL. A few 50 cycles between trajectory frames with a minimum time length of 10 ns at a temperature of 1.4 °C were applied to observe the protein motion. Other additional distance restraint parameters were set to default. Root Mean Square Fluctuation (RMSF) value was utilized to assess the stability of the protein during simulation.

Results and Discussion

A total of 13088 SNPs data for hCRY2 protein were retrieved from the NCBI dbSNP database. Out of these, 10846 SNPs were found to be present in the intron region (82.9%), 436 were nonsynonymous or missense SNPs (3.3%), and 230 were synonymous variants (1.8%). Only missense SNPs were considered for further analysis. In detail, information about all the missense variants is given in Supplementary Table 1. Six in silico tools like PolyPhen2, SNAP2, PONP2, Rhapsody, Pmut and SNPs&GO were utilized to find out the deleterious, pathogenic, and disease-causing SNPs. All prediction tools identified 68 variants out of 436 nsSNPS as disease-causing SNPs, which are listed in Supplementary Table 2. These 68 variants were then assessed for structural stability of hCRY2 upon missense mutation using MUpro, mCSM, DeepDDG, INPS-3D, and PremPS (Table 1). These tools showed the predicted changes in folding Gibbs free energy (ΔΔG). For MUpro, mCSM, DeepDDG, and INPS-3D, variants with ΔΔG value < -1.0 kcal/mol, and for PremPS ΔΔG value > 1.0 kcal/mol were considered to have significantly destabilized hCRY2 structure. Overall, fifteen missense variants that showed significant destabilization across four or more tools were considered deleterious to hCRY2 stability. In certain cases, fifteen more variants that displayed destabilization across three tools were labeled as marginally deleterious to hCRY2 stability.

Project HOPE web server was used to predict the effects of missense variants on protein by analyzing the sequences of the 16 variants selected from the previous analysis. The obtained results assessed the changes in contacts between residues within the protein, perturbation in the protein structure, impacts of the functional domains of the protein, differences of the protein, and differences in the properties of the wild type and the substituted amino acid. The analysis of each variant is displayed in Table 2. All the amino acid residues of the variants analyzed by Consurf were reported to be highly conserved. The W371R, F400I, 2F429C, V380M, W71R, A407T, F315V, I359M, L74P, L274P, Y169H, L115W, L214Q, L288Q, L309P and C433R variants were predicted to be conserved as buried and structural while the conservation of other variants was functional and exposed. The conservation profile of the variants is represented in Supplementary File 1. From the analysis, five variants, L74P, L274P, L309P, F315V, and Y485, significantly affected hCRY2 structure, stability, and binding to both FBXL3 and PER2.

Hence, the three-dimensional structures of the native and mutant CRY2 variants were generated using homology by SWISS-MODEL (Fig.2). The structures were also validated using SAVES 6.0 server (Table 3). The PROCHECK results predicted the amino acid residues of all the native and mutant models results in the most favored regions were more than 83%, indicating the good quality of the models in the Ramachandran plots of the 3D models (Supplementary Fig. 2). The ERRAT value analysis demonstrated that the nonbonded interactions and backbone confirmation of the predicted models fit well within the range of a good quality model which was above 91%.

Complexes of hCRY2 and PER2, and hCRY2 and FBXL3 were generated via SWISS-MODEL (Fig.3). This was done to assess binding free energy change (ΔΔG values) in protein-protein interactions upon missense mutation that is crucial for instability of the protein interactions, functions, and pathogenicity (Jemimah and Gromiha, 2020). Four tools, mCSM-PPI2, MutaBind2, SAAMBE3D, and BeAtMuSic V1.0, were applied to predict changes in the binding affinity of each complex (Table 4). For MutaBind2, SAAMBE3D and BeAtMuSic V1.0 ΔΔG values > 0.8 kcal/mol, and for mCSM-PPI2, ΔΔG values < -0.8 kcal/mol reflect a significant decrease in binding affinity to the binding partners. Consequently, variants that followed these criteria across three or more tools were considered to have destabilized binding with its partner protein. When bound to FBXL3, fourteen variants, and in PER2 bound state, nine variants of hCRY2 followed these criteria.

Table 1. Impacts of missense variants on hCRY2 protein stability.

Variants	MUpro	mCSM	DeepDDG	INPS3D	PremPS
W371R	-1.081	-2.161	-1.965	-2.330	1.91
F400I	-0.631	-0.428	-1.344	-2.537	1.67
P488S	-1.463	-1.707	-0.851	-1.134	1.13
R239C	-0.723	-1.946	-1.378	-1.085	1.11
F429C	-1.579	-1.106	-0.482	-2.179	1.29
V380M	-0.387	-0.666	-1.084	-1.709	1.18
R282H	-1.506	-1.998	-2.728	-1.070	1.2
R377Q	-1.357	-1.767	-1.087	-0.846	1.51
W71R	-1.667	-0.3	-2.017	-2.387	1.16
A407T	-0.529	-1.095	-1.972	0.028	1.2
G231R	-1.517	-0.82	-2.103	-0.597	1.82
F315V	-1.450	-0.253	-0.276	-2.030	1.21
R70P	-0.754	-1.124	-1.488	-1.897	1.47
R312Q	-0.183	-0.624	-0.629	-1.481	0.41
G370V	-0.826	-1.428	-1.25	-1.192	1.15
I359M	-2.079	-0.578	-4.405	-1.304	2.58
L74P	-1.329	-0.93	-1.523	-3.368	0.94
L274P	-2.432	-1.944	-4.182	-3.365	2.44
I337T	-1.946	-0.328	-2.681	-2.539	1.94
Y464C	-1.109	-2.353	-0.357	-2.039	2.14
Y485H	-1.041	-2.239	-1.596	-1.272	2.12
R70G	-1.270	-2.384	-2.235	-1.996	1.6
Y169H	-1.356	-1.162	-2.818	-1.551	1.79
N37K	-1.441	-0.594	-4.001	-0.767	1.64
L115W	-1.788	0.078	-3.371	-2.238	0.79
Y152C	-1.428	-2.189	-1.246	-1.695	2.31
L214Q	-1.289	-1.133	-2.861	-2.035	1.96
L288Q	-2.124	-1.268	-3.621	-2.406	1.83
L309P	-1.064	-0.589	-5.336	-3.469	2.55
C433R	-1.521	0.397	-0.811	-1.145	1.09

The bold letter indicates the significant destabilizing missense variants predicted by four or more tools. Numerical values in the table show the expected changes in folding Gibbs free energy (ΔΔG) in kcal/mol.

Table 2. Structural and functional impacts of hCRY2 missense variants

Variants	Variation in contact	Positional disturbance	Alteration in amino acid properties
R282H	The variant doesn't make the hydrogen bond at Asn 147 and Thr 150 and disturbs the ionic interaction.	Mutant residue might disturb the core structure of this domain.	The mutant residue is smaller. This causes charge differences and creates an empty space in the core of the protein.
W71R	The mutation might cause a loss of hydrophobic interactions with other molecules on the protein's surface.	The mutated residue is located on the surface of a domain, which can disturb this domain and abolish its function.	The mutant residue is smaller, causing repulsion between neighboring residues.
R70P	The variant doesn't make the hydrogen bond with Asp 401 and disturbs the ionic interaction.	The mutated residue is located on the surface of a domain, which can disturb this domain and abolish its function.	This mutation loses the charge of the wild-type residue. This can cause a loss of interactions with other molecules.
L74P	Variant residue disrupts the α-helix and can severely affect the structure of the protein.	The residue is buried in the core of a domain that might disturb the core structure of this domain.	The mutant residue is smaller and causes an empty space in the core of the protein.
L274P	Variant residue disrupts the α-helix and can severely affect the structure of the protein.	The residue is buried in the core of a domain that might disturb the core structure of this domain.	The mutant residue is smaller, causing a space in the core of the protein.
I337T	The mutation might cause a loss of hydrophobic interactions in the core of the protein.	The residue is buried in the core of a domain that might disturb the core structure of this domain.	The mutant residue is smaller, causing an empty space in the core of the protein.
Y485H	The variant doesn't make the hydrogen bond with Asp 360 and loss of hydrophobic interactions in the core of the protein.	The mutation is in a region required for the inhibition of CLOCK-BMAL1-mediated transcription. The residue is buried in the core of a domain that might disturb the core structure.	The mutant residue is smaller, causing an empty space in the core of the protein.
R70G	The variant doesn't make the hydrogen bond and salt bridge with 401, disturbs ionic interactions, and causes a loss of hydrophobic interactions in the core of the protein. It also interferes with the rigidity of the protein.	The mutated residue is located on the surface of the Photolyase/cryptochrome alpha/beta domain, which can disturb this domain and abolish its function.	The mutant residue is smaller.
Y169H	The variant doesn't make the hydrogen bond with Gln 173 and loss of hydrophobic interactions on the protein's surface.	The variant residue is located on the surface of a domain with an unknown function.	The mutant residue is smaller.
Y152C	The mutation may cause a loss of hydrogen bonds in the core of the protein and, as a result, disturb correct folding.	The residue is buried in the core of a domain located in a β-strand. The variant residue may slightly destabilize local conformation.	The mutant residue is smaller than the wild-type residue, causing an empty space in the core of the protein.
L214Q	The variant residue may cause a loss of hydrophobic interactions in the core of the protein.	The residue is buried in the core of a domain that might disturb the core structure of this domain.	The mutant residue is bigger.
L288Q	The variant residue may cause a loss of hydrophobic interactions in the core of the protein.	The residue is buried in the core of a domain that might disturb the core structure of this domain.	The mutant residue is bigger.
L309P	Variant residue disrupts the α-helix and can severely affect the structure of the protein and ligand contacts and destabilize the stability.	The residue is buried in the core of a domain that might disturb the core structure, affect the local structure, and consequently affect this binding site.	The mutant residue is smaller, causing an empty space in the core of the protein.
F315V	The variant residue may cause a loss of interactions with ligands and interfere with the protein's function.	The residue is buried in the core of a domain that might disturb the core structure of this domain.	The mutant residue is smaller, causing an empty space in the core of the protein.

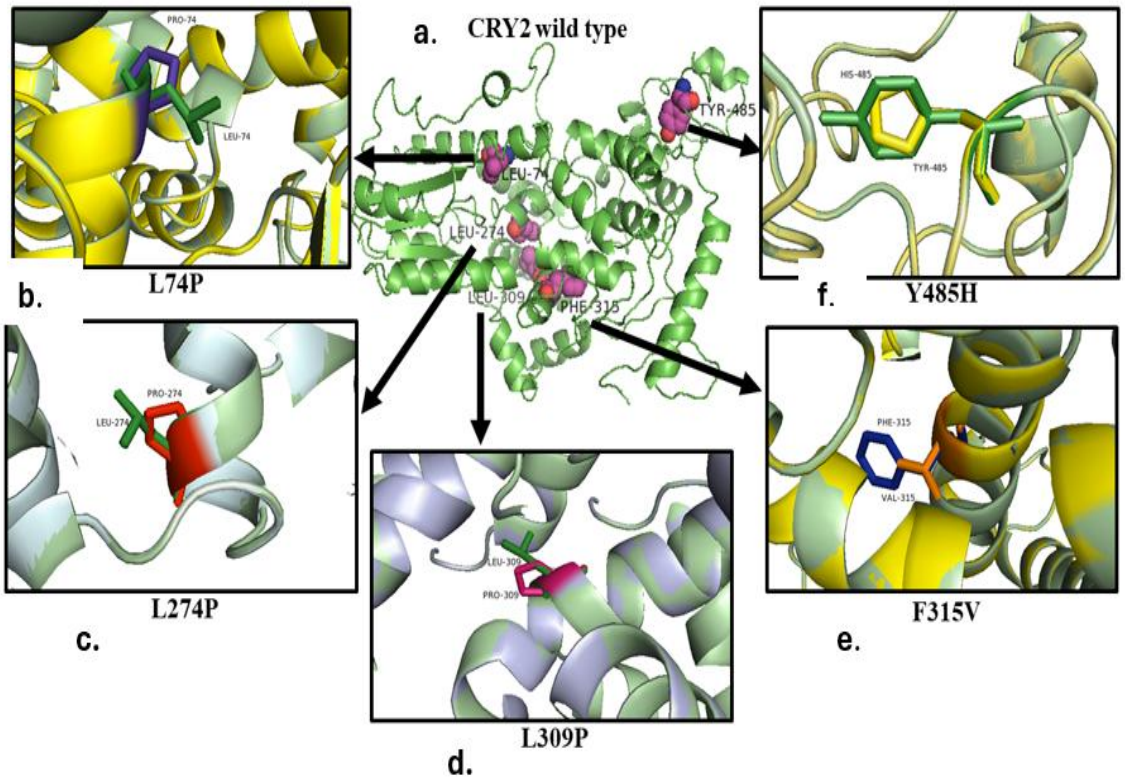

Fig. 2. 3D structure of native and mutant hCRY2 variants a. native hCRY2 b. L74P c. L274P d. L309P e. F315V f. Y485H.

Table 3. Validation scores of predicted models obtained from PROCHECK and ERRAT.

Variant	PROCHECK (%)[a]	ERRAT(%)[b]
Wild type (hCRY2)	85.6	91.111
L74P	86.9	93.915
L274P	87.1	94.118
L309P	85.6	91.111
F315V	85.6	91.111
Y485H	85.6	90.303

[a]residues in most allowed regions of Ramachandran plot [b]Overall Quality Factor expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit with structure resolutions (2.5 to 3A).

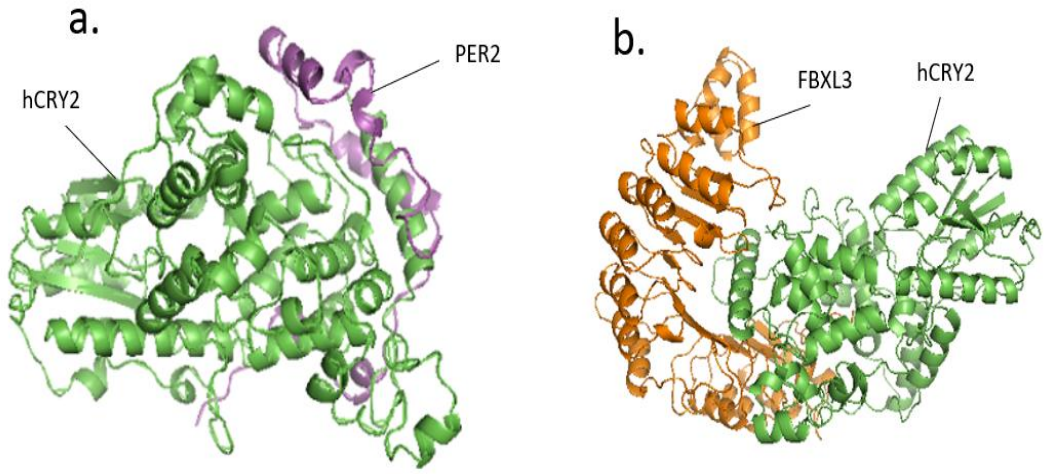

Fig. 3. (a) Tertiary structure of PER2-hCRY2 complex (b) 3D structure of FBXL3-hCRY2 complex. hCRY2, PER2, and FBXL3 are denoted by the green, magenta-colored, and orange-colored ribbon

Table 4. Effects of missense SNPs on protein-protein interactions between hCRY2 and FBXL3 and hCRY2 and PER2.

hCRY2 and FBXL3 complex					hCRY2 and PER2 complex				
Variants	**mCSM-PPI**	**MutaBind2**	**SAAM BE3D**	**BeAtMusicV1.0**	**Variants**	**mCSM-PPI**	**Muta Bind2**	**SAAM BE3D**	**BeAtMusicV1.0**
R440H	-0.416	1.44	1.09	0.82	R171C	-0.687	1.88	0.86	1.31
R440C	-0.895	1.16	0.8	1.52	W371R	-0.202	1.12	1.4	0.81
F429C	-1.95	2.98	2.29	2.38	F429C	-0.923	0.89	0.92	0.87
R377Q	-1.621	1.31	1.31	0.28	G370V	-1.025	0.98	0.14	0.82
R377L	-1.35	2.38	0.95	-0.29	**L74P**	-0.931	0.68	0.81	1.21
F315V	-1.574	1.05	0.81	0.71	**L274P**	-0.567	0.91	0.97	1.25
R312Q	-1.681	1.25	1.02	0.77	**Y485H**	-1.168	0.75	0.98	0.93
G370V	-0.847	1.21	0.14	1.05	W416C	-0.542	1.01	1.26	1.07
L74P	-0.937	1.23	0.81	1.22	**L309P**	-0.763	0.86	1.14	1.54
L274P	-1.572	1.44	0.97	2.08					
Y169H	-1.326	0.84	0.94	1.27					
H374P	-1.429	1.1	0.77	.81					
W416C	-0.842	1.11	1.51	1.19					
L309P	-1.405	1.08	1.41	1.89					

The bold letter indicates the significant destabilizing missense variants predicted by four tools.

These five variants, L74P, L274P, L309P, F315V, and Y485H, were thus selected for protein dynamics simulation to determine the protein flexibility by calculating the RMSF of all amino acids and the contact map between residues of the CRY2 protein. The resulting residue fluctuation profiles (Fig.4) show specific changes compared to the wild type. The difference of1Å in the RMSF value of wild-type and that of mutant hCRY2 variants was residues that may form unstable secondary and tertiary structures. In all 5 variants, the residues showed prominent fluctuations in the DNA photolyase region (22-176) and the region that inhibits the CLOCK-ARNTL-driven transcription (390-489) as compared to the native protein. From the analysis, it can be observed that Leu274 residue is placed in the FAD binding domain. Upon alteration to a 5-membered proline which resides in the buried area of the protein destabilizes the structure by the disruption of α-helix and creating a space in the protein. Previous experimental studies found that point mutations at amino acid residues crucial for the FAD binding domain (246-444) affect the repressor activity of CRY2 (Czarna et al., 2013). Hirano et al. (2016) identified a missense variant (A260T) in the FAD binding domain of the human hCry2 gene that leads to heritable Familial Advanced Sleep Phase (FASP)-like phenotype. Moreover, conversion from serine to aspartic acid at the 265th position of mouse CRY2 (homologous to Ser266 in hCRY2) that resides in the phosphorylation site declines CRY2 repressor activity (Sanada et al., 2004). L274P alteration also remarkably decreases the binding affinity and interferes with the interactions of FBXL3 and PER2 complex that are essential for the regulation of CRY2 stability and maintenance of the sleep-wake cycle in humans (Hirano et al., 2016).

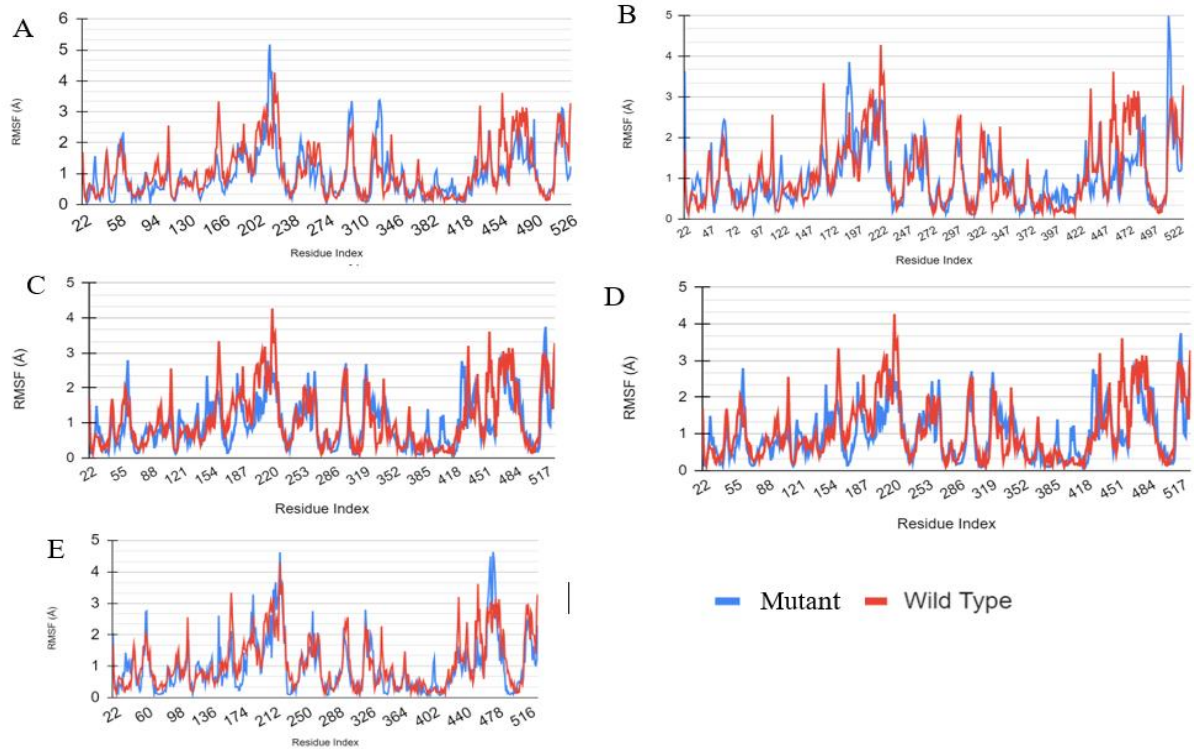

Fig. 4. Residue fluctuation profiles of the native and five mutant structures of hCRY2. The variants are (A) L74P, (B) L274P, (C) L309P, (D) F315V, (E)Y485H. All these variants were predicted to alter the wild-type protein structure and reduce hCRY2 stability.

The Glu308 residue is considered the FAD binding site in human CRY2 protein. The alteration of the next residue, leucine to proline that is, evolutionarily conserved residue, destabilizes protein structure by alteration of α-helix, ligand-contacts, and binding sites and makes a space in the core of the protein. The present study found that mutation from leucine to proline at 309th position significantly reduces the binding affinity of FBXL3 to hCRY2 protein. In the case of mCRY2, Trp428 residue of the C-terminal tail of FBXL3 binds to the center of the FAD-binding pocket of mCRY2 and occupies the cofactor site. The tail of FBXL3 loads its Pro426 residue against Trp310 of mCRY2 (homologous to Trp311 in hCRY2) and turns the side chain His373 with Thr427 of mCRY2, to enter the FAD-binding pocket. Inside the pocket, the indole ring of Trp428 forms several hydrophobic interactions with mCRY2 and donates a hydrogen bond to the carbonyl group of Gln307 (homologous to Gln308 in hCRY2) in mCRY2 (Xing et al., 2013). FAD and FBXL3 competitively bind to the same FAD binding pocket and promote ubiquitylation of CRY proteins (Hirano et al., 2017). So, it can be predicted that hydrogen and hydrophobic interactions necessary for binding FBXL3 to the core of the FAD binding pocket may be hampered by substituting nonpolar amino acid leucine for polar amino acid proline. MD simulation study also suggests that the mutation alters the position of amino acids at several points of photolyase region (162-163,175-176) and CLOCK-ARNTL-mediated transcription inhibition region (433-435,454-458 and 478-83) that may lead to perturbation of the circadian clock.

The FAD binding pocket comprises three different sub-regions: hydrophobic region 1, affinity region, and hydrophobic region 2. Residue W310, F314, W417 and L418 is located in the hydrophobic region 1 of mCRY2 (homologous to W311, F315, W418, and L418 in hCRY2) (Miller et al., 2021). Minimal changes in the hydrophobic region significantly hamper the binding modes of CRY activating compounds, cyclopentyl group of TH301, chlorobenzonitrile moiety of KL044, and the more flexible methane sulfonamide and furan groups of KL001 of mCRY2 protein (Miller and Hirota, 2022). The study reported that mutation from aromatic amino acid phenylalanine to aliphatic amino acid valine at the 315th position may cause loss of interactions with ligands, destabilize the core structure of the domain, and interfere with protein's function. So, the alteration may cause the structural differences of FAD binding pockets that change the binding mode of other compounds as well as reduce the binding affinity to FBXL3 and PER2 protein. When simulated, the mutation fluctuates the photolyase homology region (163-165) and CLOCK-ARNTL-mediated transcription inhibition region (454-462, 478-483), which may imbalance the CRY2 degradation and stabilization.

The wild-type tyrosine residue lies in the region required to inhibit CLOCK-ARNTL-mediated transcription, which is the central negative feedback loop of the circadian clock (Partch et al., 2014; Xing et al., 2013). Alteration of tyrosine residues to histidine at the 485th position may change hydrogen and hydrophobic interactions in the core of the protein and disturb the domain structure. The variation also decreases the interactions of CRY2 protein with PER2 and FBXL3, which may hamper the rhythm of the circadian clock.

The impacts of dysfunction of human CRY2 protein are associated with multiple abnormalities. The pathophysiology of cardiovascular disease, of which hypertension is a significant contributing element, has been linked to circadian clock disruption (Škrlec et al., 2018). It appears to impact not only the cycles of metabolism and cell division but also mood and behavior (Kovanen et al., 2014), depression (particularly winter depression), bipolar type 1, and seasonal affective disorder (Kovanen et al., 2013; Lavebratt et al., 2010b). On the other hand, tumorigenesis might be aided by the disturbance of circadian rhythms. Thus, various features of carcinogenesis, such as cell growth, angiogenesis, metabolism, apoptosis, and DNA damage response, may be linked to circadian negative feedback loop

(CNFL) genes (Qu et al., 2016). Indeed, a variety of human cancer types commonly exhibit altered expression of the CNFL genes (CRY1, CRY2, PER1, PER2, and PER3) (Gršković & Korać, 2023).

Conclusion

Dysfunction of the human CRY2 protein is associated with multiple abnormalities. Non-synonymous polymorphisms of the CRY2 gene result in an altered protein that may interfere with the morning/evening preference and may disturb different metabolic regulation and circadian periods. This study determined L74P, L274P, L309P, F315V, and Y485H to be the most deleterious missense mutants in the hCRY2 protein. These variants are predicted to affect the binding of FAD cofactors, change the binding pocket structure, and alter the binding mode of different CRY-activating compounds. These mutants may destabilize circadian period length regulation by reducing the binding affinity to FBXL3 and PER2. The most deleterious mutants found can undergo in vitro assays to understand protein-protein interactions of these mutants in cell culture, Xray crystallography, and Mass Spectrometry to confirm the changes in structures of these proteins produced in cell culture, and many other experimental procedures. Further experimental studies are needed to validate the effects of these SNPs in animal models, like in mutant and wild-type mice, by observing their behavioral and genetic effects and protein expression levels that may lead to establishing the real impacts of SNPs on the circadian clock of mice.

Author contributions

ASK took part in acquiring, analyzing, and interpreting data and manuscript preparation. MA participated in the study and interpretation of data, manuscript preparation, and review. MAE was responsible for data acquisition and analysis. SFK was involved in the conception, analysis, interpretation of data, and critical reviewing. All authors have read and approved the final manuscript.

Declaration of conflicting interests

The authors declare that they have no competing interests.

Supplementary materials

The single nucleotide polymorphism datasets analyzed are available in the dbSNP repository (Accession Number: NP_066940.3). The amino acid sequence of hCRY2 used to generate the wild-type structure was collected from the UniProt Knowledge Base (UniProtKB) (Accession Number: Q49AN0). All other data generated or analyzed during this study are included in this article (and its supplementary information files).

References

Adzhubei I, Jordan DM and Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet*. 2013; 76(1): 7.20.21-27.20.41.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T and Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016; 44(W1): W344-W350.

Buhr ED and Takahashi JS. Molecular components of the mammalian circadian clock. *Handb. Exp. Pharmacol.* 2013; 217, 3-27.

Calabrese R, Capriotti E, Fariselli P, Martelli PL and Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat*. 2009; 30(8): 1237-1244.

Cao H, Wang J, He L, Qi Y and Zhang JZ. DeepDDG: predicting the stability change of protein point mutations using neural networks. *J. Chem. Inf. Model.* 2019; 59(4): 1508-1514.

Chan AB, Huber AL and Lamia KA. Cryptochromes modulate E2F family transcription factors. *Sci. Rep*. 2020; 10(1): 1-9.

Chasman D and Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-

based assessment of amino acid variation. *J. Mol. Biol.* 2001; 307(2): 683-706.

Cheng J, Randall A and Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins: Struct., Funct., and Bioinf. 2006; 62(4): 1125-1132.

Colovos C and Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein sci.: a publication of the Protein Society. 1993; 2(9): 1511-1519.

Correia SP, Chan AB, Vaughan M, Zolboot N, Perea V, Huber AL, Kriebs A, Moresco JJ, Yates JR and Lamia K A. The circadian E3 ligase complex SCF FBXL3+CRY targets TLK2. *Sci. Rep*. 2019; 9(1).

Czarna A, Berndt A, Singh HR, Grudziecki A, Ladurner AG, Timinszky G, Kramer A and Wolf E. Structures of Drosophila cryptochrome and mouse cryptochrome1 provide insight into circadian function. *Cell.* 2013; 153(6): 1394-1394.

Dehouck Y, Kwasigroch, JM, Rooman M and Gilis D. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res.* 2013; 41(W1): W333–W339.

Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson A U, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Mägi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V and Barroso I. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet*. 2010; 42(2): 105-116.

Fribourgh JL, Srivastava A, Sandate CR, Michael AK, Hsu PL, Rakers C, Nguyen LT, Torgrimson MR, Parico GCG, Tripathi S, Zheng N, Lande GC, Hirota T, Tama F and Partch CL. Dynamics at the serine loop underlie differential affinity of cryptochromes for CLOCK:BMAL1 to control circadian timing. *eLife*. 2020; 9.

Gekakis N, Staknis D, Nguyen HB, Davis FC, Wilsbacner LD, King DP, Takahashi JS and Weitz CJ. Role of the CLOCK protein in the mammalian circadian mechanism. *Science,* 1998; 280(5369): 1564-1569.

Gršković P and Korać P. Circadian Gene Variants in Diseases. *Genes,* 2023; 14(9): 1703-1703.

Hecht M, Bromberg Y and Rost B. Better prediction of functional effects for sequence variants. *BMC Genom.* 2015; 16 (8): S1.

Hirano A, Braas D, Fu YH and Ptáček LJ. FAD regulates CRYPTOCHROME protein stability and circadian clock in mice. *Cell Rep.* 2017; 19(2): 255-266.

Hirano A, Shi G, Jones CR, Lipzen A, Pennacchio LA, Xu Y, Hallows WC, McMahon T, Yamazaki M, Ptáček LJ and Fu YH. A Cryptochrome 2 mutation yields advanced sleep phase in humans. *eLife*. 2016; 5: e16695.

Hoffman AE, Zheng T, Stevens RG, Ba Y, Zhang Y, Leaderer D, Yi C, Holford, TR and Zhu Y. Clock-cancer connection in non-Hodgkin's lymphoma: A genetic association study and pathway analysis of the Circadian gene Cryptochrome 2. *Cancer Res.* 2009; 69(8): 3605-3613.

Hogenesch JB, Gu YZ, Jain S and Bradfield CA. The basic-helix–loop–helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proc. Natl. Acad. Sci. USA*, 1998; 95(10): 5474-5474.

Huber AL, Papp SJ, Chan AB, Henriksson E, Jordan SD, Kriebs A, Nguyen M, Wallace M, Li Z, Metallo CM and Lamia KA. CRY2 and FBXL3 Cooperatively Degrade c-MYC. *Mol. Cell*, 2016; 64(4): 774-789.

Jemimah S and Gromiha MM. Insights into changes in binding affinity caused by disease mutations in protein-protein complexes. *Comput. Biol. Med.* 2020; 123: 103829.

Kelly MA, Rees SD, Hydriem ZL, Shera AS, Bellary S, O'Hare JP, Kumar S, Taheri S, Basit A and Barnett AH. Circadian gene variants and susceptibility to type 2 diabetes: A pilot study. PLoS One. 2012; 7(4):e32670.

King DP, Zhao Y, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch, MP, Steeves, TDL, Vitaterna MH, Kornhauser JM, Lowrey PL, Turek FW and Takahashi JS. Positional cloning of the mouse circadian clock gene. Cell, 1997; 89(4): 641-653.

Kovanen L, Donner K, Kaunisto M and Partonen T. CRY1, CRY2 and PRKCDBP genetic variants in metabolic syndrome. *Hypertension Res*. 2014; 38(3): 186-192.

Kovanen L, Kaunisto M, Donner K, Saarikoski ST and Partonen T. CRY2 genetic variants associate with dysthymia. *PLoS One*, 2013; 8(8): e71450.

Kripke DF, Nievergelt CM, Joo EJ, Shekhtman T and Kelsoe JR. Circadian polymorphisms associated with affective disorders. *J. Circadian Rhythms*, 2009; 7: 2.

Kucukkal TG, Petukh M, Li L and Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr. Opin. Struct. Biol.* 2015; 32: 18-24.

Kumar A and Purohit R. Computational screening and molecular dynamics simulation of disease associated nsSNPs in CENP-E. *Mutat. Res.* 2012; 738-739: 28-37.

Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M and Kmiecik S. CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures. *Nucleic Acids Res*. 2018; 46(W1): W338-W343.

Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R and Thornton, JM. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR.* 1996; 8(4): 477-486.

Lavebratt C, Sjöholm LK, Soronen P, Paunio T, Vawter MP, Bunney WE, Adolfsson R, Forsell Y, Wu JC, Kelsoe, JR, Partonen T and Schalling M. CRY2 Is Associated with Depression. *PLoS One*. 2010(a); 5(2): e9407.

Lavebratt C, Sjöholm LK, Soronen P, Paunio T, Vawter MP, Bunney, WE, Adolfsson R, Forsell Y, Wu JC, Kelsoe, JR, Partonen, T and Schalling M. CRY2 is associated with depression. *PLoS One.* 2010b; 5(2): e9407.

Li Y, Zhou J, Wu Y, Lu T, Yuan M, Cui Y, Zhou Y, Yang, G and Hong, Y. Association of osteoporosis with genetic variants of circadian genes in Chinese geriatrics. *Osteoporos. Int*. 2016; 27(4): 1485-1492.

López-Ferrando V, Gazzo A, De La Cruz X, Orozco M and Gelpí JL. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res*. 2017; 45(W1): W222-W228.

Miller S and Hirota T. Structural and chemical biology approaches reveal isoform-selective mechanisms of ligand interactions in mammalian cryptochromes. *Front. Physiol.* 2022; 13: 837280.

Miller S, Srivastava A, Nagai Y, Aikawa Y, Tama F and Hirota T. Structural differences in the FAD-binding pockets and lid loops of mammalian CRY1 and CRY2 for isoform-selective regulation. *Proc. Natl. Acad. Sci. USA*, 2021; 118(26): e2026191118.

Mohawk JA, Green CB and Takahashi JS. Central and peripheral circadian clocks in mammals. *Annu. Rev. Neurosci.* 2012. 35: 445-462.

Nangle SN, Rosensweig C, Koike N, Tei H, Takahashi JS, Green CB and Zheng N. (2014). Molecular assembly of the period-cryptochrome circadian transcriptional repressor complex. *eLife,* 2014. 3: e03674.

Niroula A, Urolagin S and Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS One, 2015; 10(2): e0117380.

Pahari S, Li G, Murthy, AK, Liang S, Fragoza R, Yu H and Alexov, E. (2020). SAAMBE-3D: Predicting Effect of Mutations on Protein–Protein Interactions. Int. J. Mol. Sci. 2020; 21(7): 2563.

Partch CL and Sancar A. Photochemistry and photobiology of cryptochrome blue-light

photopigments: The search for a photocycle. *Photochem. Photobiol.* 2005; 81(6): 1291-1304.

Partch CL, Green CB and Takahashi JS. Molecular architecture of the mammalian circadian clock. *Trends Cell Biol*. 2014; 24(2): 90-99.

Petukh M, Kucukkal TG and Alexov E. On human disease-causing amino acid variants: Statistical Study of sequence and structural patterns. *Hum. Mut.* 2015; 36(5): 524-534.

Pires DE, Ascher DB and Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinf.* 2014; 30(3): 335-342.

Ponzoni L, Peñaherrera DA, Oltvai ZN and Bahar I. Rhapsody: predicting the pathogenicity of human missense variants. *Bioinf.* 2020; 36(10): 3084-3092.

Qu F, Qiao Q, Wang N, Ji G, Zhao H, He L, Wang H and Bao G. Genetic polymorphisms in circadian negative feedback regulation genes predict overall survival and response to chemotherapy in gastric cancer patients. *Sci. Rep.* 2016; 6: 22424.

Rajendran V and Sethumadhavan R. Drug resistance mechanism of PncA in Mycobacterium tuberculosis. *J. Biomol. Struct. Dyn.* 2014; 32(2): 209-221.

Rajendran V, Gopalakrishnan C and Sethumadhavan R. Pathological role of a point mutation (T315I) in BCR-ABL1 protein—A computational insight. *J. Cell. Biochem.* 2018 119(1), 918-925.

Reischl S and Kramer A. Kinases and phosphatases in the mammalian circadian clock. *FEBS Lett*. 2011; 585(10): 1393-1399.

Reppert SM and Weaver, DR. Coordination of circadian timing in mammals. *Nature,* 2002; 418(6901): 935-941.

Rodrigues CHM, Myung Y, Pires DEV and Ascher DB. mCSM-PPI2: predicting the effects of mutations on protein–protein interactions. *Nucleic Acids Res*. 2019; 47(W1): W338-W344.

Rosensweig C, Reynolds, KA, Gao P, Laothamatas I, Shan, Y, Ranganathan R, Takahashi JS and Green CB. An evolutionary hotspot defines functional differences between CRYPTOCHROMES. *Nat. Commun.* 2018; 9(1): 1138.

Salazar P, Konda S, Sridhar A, Arbieva Z, Daviglus M, Darbar D and Rehman J. Common genetic variation in circadian clock genes are associated with cardiovascular risk factors in an African American and Hispanic/Latino cohort. *Int. J. Cardiol. Heart Vascul.* 2021; 34: 100808.

Sanada K, Harada Y, Sakai M, Todo T and Fukada Y. Serine phosphorylation of mCRY1 and mCRY2 by mitogen-activated protein kinase. *Genes Cells.* 2004; 9(8): 697-708.

Savojardo C, Fariselli P, Martelli PL and Casadio R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinf.* 2016; 32(16), 2542-2544.

Schork NJ, Fallin D and Lanchbury JS. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin. Genet.* 2000; 58(4): 250-264.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski, EM and Sirotkin, K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*2001; 29(1): 308-311.

Škrlec I, Milic J, Heffer M, Peterlin, B and Wagner J. Genetic variations in circadian rhythm genes and susceptibility for myocardial infarction. *Genet. Mol. Biol.* 2018; 41(2): 403-409.

Stojkovic K, Wing SS and Cermakian N. A central role for ubiquitination within a circadian clock protein modification code. Front. Mol. Neurosci. 2014; 7: 69.

Takahashi JS. Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* 2016; 18(3): 164-179.

Venselaar H, te Beek TAH, Kuipers RKP, Hekkelman ML and Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinf.* 2010; 11: 548.

Vielhaber EL, Duricka D, Ullman KS and Virshup DM. Nuclear export of mammalian PERIOD proteins. *J. Biol. Chem*. 2001; 276(49): 45921-45927.

Wang Y, Peng S, Liu T, Zhang Y, Li H, Li X, Tao W and Shi Y. The potential role of clock genes in children attention-deficit/hyperactivity disorder. *Sleep Med.* 2020; 71: 18-27.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T, de Beer TAP, Rempfer C, Bordoli L, Lepore R and Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018; 46(W1): W296-W303.

Xing W, Busino L, Hinds TR, Marionni ST, Saifee NH, Bush MF, Pagano M and Zheng N. SCFFBXL3 ubiquitin ligase targets cryptochromes at their cofactor pocket. *Nature.* 2013; 496(7443): 64-68.

Zhang N, Chen Y, Lu H, Zhao F, Alvarez, RV, Goncearenco A, Panchenko AR and Li M. MutaBind2: Predicting the impacts of single and multiple mutations on protein-protein interactions. *iScience*. 2020; 23(3): 100939.

Zienolddiny S, Haugen A, Lie JAS, Kjuus, H, Anmarkrud KH and Kjærheim, K. Analysis of polymorphisms in the circadian-related genes and breast cancer risk in Norwegian nurses working night shifts. *Breast Cancer Res.* 2013; 15(4): R53.