ISTANBUL UNIVERSITY PRESS

RESEARCH ARTICLE

Eur J Biol 2022; 81(2): 125-135

In Silico Analysis of BMAL1 and CLOCK SNPs in the Ensembl Database

Seref Gul[1]

[1]Istanbul University, Faculty of Science, Department of Biology, Biotechnology Division, Istanbul, Turkiye

ORCID IDs of the authors: S.G. 0000-0002-5613-1339

Please cite this article as: Gul S. In Silico Analysis of BMAL1 and CLOCK SNPs in the Ensembl Database. Eur J Biol 2022; 81(2): 125-135. DOI: 10.26650/EurJBiol.2022.1164864

ABSTRACT

Objective: A circadian rhythm in mammals controls the sleep-wake cycle, blood pressure, hormone secretion, metabolism and many other physiological processes. The circadian clock mechanism is regulated by four genes: *Bmal1, Clock, Cry,* and *Per.* Mutations in these regulatory genes are associated with sleep and mood disorders, obesity, and cancer. Several PER2 and CRY2 SNPs are associated with advanced sleep phase syndrome. It is, therefore, critical to understand the effect of clock genes' SNPs on the circadian clock. In this study, we determined "pathogenic" BMAL1 and CLOCK SNPs in the Ensembl database for biochemical characterization.

Materials and Methods: BMAL1 and CLOCK SNPs in the Ensemble database were filtered out for only missense mutations. Among the missense mutations, pathogenic ones were determined according to SIFT, PolyPhen, and CADD scores, REVEL, MetaIR, Mutation Assessor, I-Mutant, PROVEAN, and FireDock programs. BMAL1 and CLOCK SNPs were visualized by using PyMol.

Results: Thousands of BMAL1 and CLOCK missense SNP mutations were reported in the Ensembl database. After the classification of those SNPs according to their SIFT, PolyPhen, and CADD pathogenicity, twelve SNPs for each protein remained as pathogenic. A further analysis with all in silico tools revealed that BMAL1 SNPs causing Ala154Val, Arg166Gln, and Val440Gly mutations; and CLOCK SNPs causing Gly120Val, Asp119Val, Gly120Ser, Ala117Val, and Cys371Gly mutations were predicted as the most "pathogenic" ones.

Conclusion: Overall, by using in silico tools, we provided a starting point for experimental studies for determining the effect of pathogenic BMAL1 and CLOCK SNPs on the circadian clock mechanism.

Keywords: Circadian clock, Ensembl, BMAL1, CLOCK, SNP

INTRODUCTION

The circadian clock is an endogenous biological system that generates 24-hour rhythmic changes in the physiology and behavior of organisms according to environmental factors such as light-dark and eating-fasting cycles or social interactions. The circadian clock mechanism is found in a wide variety of organisms from cyanobacteria to drosophila to mammals (1). The circadian clock determines the correct timing of various biochemical processes in mammals such as the sleep-wake cycle, metabolism, body temperature, hormone secretion, detoxification, sexual functions, DNA repair and apoptosis (2-5). Factors such as the night-day cycle, mealtime, social interaction, emotional mood, and temperature synchronize the circadian clock to exactly 24 hours (6,7). Among those, light turned out to be the most dominant *zeitgeber* (time-giver) (8).

Throughout the body, each tissue has a clock known as the peripheral clock; however, the master clock located in the suprachiasmatic nucleus synchronizes all peripheral clocks to generate a unified gene transcription oscillation around a day, 24 hours (9,10). At the cellular

Corresponding Author: Seref Gul E-mail: serefgul@istanbul.edu.tr

Submitted: 21.08.2022 • **Revision Requested:** 26.09.2022 • **Last Revision Received:** 05.10.2022 • **Accepted:** 20.10.2022 • **Published Online:** 23.11.2022

level, the circadian clock mechanism is regulated by intertwined transcriptional-translational feedback loops (TTFLs). In mammals, the positive arm of the feedback loop is composed of two transcriptional factors *Bmal1* (Brain-muscle-arnt-like protein 1) and *Clock* (Circadian Locomotor Output Cycles Kaput) (11-13). The negative feedback loop is generated by the *Cry* (Cry1, Cry2) and *Per* (Per1, Per2, Per3) genes in which their transcriptions are under the control of the BMAL1 and CLOCK activity (14-16). BMAL1 and CLOCK form a heterodimer and bind to E-box sequences found in the promoter regions of the target genes and initiate their transcription (17). After *Cry* and *Per* are translated, they form a heterodimer at the cytoplasm, then translocate to the nucleus and repress the BMAL1/CLOCK dependent transcription (9,16). This mechanism controls nearly 40% of protein-coding genes in mice (18). Post-translational modifications play important roles to regulate the activity, stability, half-life, and subcellular localization of these core clock proteins (19). An auxiliary TTFL helps to maintain precise clock ticking. The retinoic acid receptor-related orphan receptors (RORs) and REV-ERBs whose transcriptions are under the control of the circadian clock bind to the ROR enhancer element to initiate and repress the *Bmal1* transcription, respectively (20,21).

When the circadian clock mechanism is disturbed or misaligned with the light-dark cycle, organisms have difficulty coping with the changes in the environment. Organisms, therefore, become more prone to diseases or syndromes. For example, cancer, depression, obesity, and some metabolic diseases are associated with the disrupted circadian clock function (5,22,23). More interestingly, some single nucleotide polymorphisms (SNPs) found in the coding region of core clock genes that change the amino acid are linked to diseases. A *PER2* SNP (p.Ser662Gly), located on a phosphorylation site, was detected in people suffering from advanced sleep phase syndrome (ASPS) (24). Another SNP, found in CRY2 (rs201220841) changing p.Ala260Thr, caused better interaction with CRY2's E3 ligase that led to its rapid degradation, was associated as well with ASPS (24). Two CRY1 exon skipping variants (CRY1 without exon 11 and exon 6) detected in the population were associated with familial delayed phase sleep syndrome and attention deficit hyperactivity disorder, respectively (25,26). Molecular characterizations showed that the exon 6 of CRY1 encodes amino acids near the CLOCK binding region, hence the absence of amino acids around this region abolished CRY1-CLOCK interaction, and mutant CRY1 (without exon 6) could not rescue the circadian rhythm at the cellular level (26). Similar characterization for CRY1 without exon 11 showed that mutant CRY1 had a higher affinity to CLOCK and BMAL1, and lengthened the period of the circadian rhythm (25). In addition to understanding the role of SNPs in human disorders, their molecular characterization may contribute to the understanding of the circadian clock mechanism better. In one of our recent studies, the characterization of a rare *CRY1* SNP (p.Arg293His) showed that two pockets of CRY1 were allosterically regulated (27). Another recent study, that analyzed rare *CRY2* SNPs, showed that amino acids around the secondary pocket are important for nuclear localization and its stability (28). It is, therefore, essential to understand how SNPs have an impact on the molecular circadian clock mechanism.

CLOCK and BMAL1 are composed of basic helix-loop-helix (bHLH) and PER-ARNT-SIM (PAS) domains. Co-crystal of BMAL1-CLOCK showed that PAS-A and PAS-B domains of these proteins interact with each other to form a heterodimer (29). PAS domains are mainly interacting with each other via hydrophobic amino acids. Leu159, Thr285, Tyr287, Val315 and Ile317 on PAS-A of BMAL1 interact with Phe104, Leu105 and Leu113 of CLOCK. PAS-B domains bind to each other in a parallel mode. Additionally, Phe423, Trp427 and Val 435 on PAS-B of BMAL1 are found to be critical for interacting with CLOCK (29). In both the BMAL1 and CLOCK, two flexible linker loops connect PAS-A and PAS-B domains. Positively charged bHLH domains of BMAL1 and CLOCK interact with each other and then are bound to the negatively charged DNA backbone (Figure 1) (30). The bHLH domains have critical hydrophobic core amino acids. For

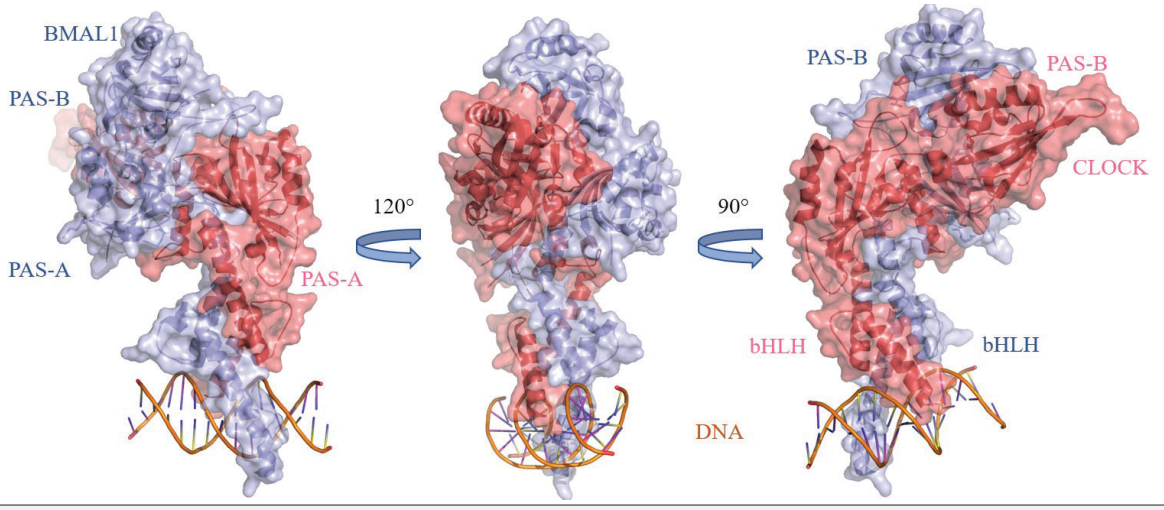


Figure 1. Superimposed crystal structures of BMAL-CLOCK-E-Box complex (PDB id: 4f3l and 4h10). BMAL1 is shown in blue, CLOCK is shown in red. Backbone of DNA is shown in orange.

example, Leu57, Leu74 of BMAL1 and Leu95, Leu115 of CLOCK were determined as critical hydrophobic amino acids in which mutating to glutamate diminished the transactivation property of BMAL1/CLOCK. In addition to the transactivation, these core hydrophobic amino acids are critical for heterodimer formation (29).

BMAL1 and CLOCK are two main transcriptional factors driving the circadian clock, therefore, determining and understanding the effect of these two proteins' SNPs in the population will advance the relationship between the circadian clock and disease; and help elucidate the molecular mechanism details of the circadian clock. While core clock genes have some paralogs (i.e., *Cry1* and *Cry2; Per1-Per2-Per3; Clock* and *Npas*), *Bmal1* has no such paralog. It is, therefore, expected that any disruption in the BMAL1 function or stability may disrupt the clock mechanism. So, the characterization of BMAL1 SNPs is of great importance. In this report, we analyzed thousands of SNPs deposited in the Ensembl database by using multiple computational tools. According to the pathogenicity criteria of various in silico analyses in the Ensembl, we ended up with twelve CLOCK and twelve BMAL1 SNPs. We expanded our calculations by using other computational tools and determined the most pathogenic SNPs by giving them pathogenicity scores. Although in silico analyses, per se, are not enough to conclude the effect of SNPs on any biological system, they provide a rational perspective for choosing a manageable number of SNPs for wet-lab experiments. By reporting pathogenic CLOCK and BMAL1 SNPs, we provided a starting point for experimental studies to analyze the effect of deleterious CLOCK and BMAL1 SNPs on the circadian clock mechanism.

MATERIALS AND METHODS

SNPs studied in this manuscript were retrieved from the Ensembl databank. Explanation of various in silico methods used to determine the effect of SNPs on BMAL1, CLOCK or BMAL1/CLOCK heterodimer are given below.

SIFT uses multiple sequence alignment approaches to determine the effect of mutations on a protein sequence (31). SIFT works in the following way: i) similar sequences are searched, ii) sequences that can show similar functions to the query sequence are selected, iii) multiple sequence alignments among selected sequences are determined, iv) normalized amino acid probability for each amino acid position is calculated. If the probability of normalized tolerability is less than 0.05, the mutation is classified as "Deleterious", if the probability is more than 0.05, it is classified as "Benign."

PolyPhen (version 2) predicts the effect of a mutation on the function of the protein by using empirical rules (32). Values larger than 0.908 were evaluated as "Probably Damaging," and values 0.446-0.908 were evaluated as "Possibly Damaging".

CADD integrates multiple parameters to one by comparing naturally selected variants and simulated mutations. The C-scores were related to the allelic variations, the pathogenicity of coding and non-coding variants, the regulating effects are determined experimentally, and a high level of pathogenic variants among individual genome sequences. CADD qualitatively prioritize functional, benign, and deleterious mutations among various functional categories, and the impact and genetic architectures (33). CADD scores larger than 30 are evaluated as "Deleterious".

REVEL is an ensemble method used to predict the pathogenicity of mutation that integrates MutPred, FATHMM, VEST, PolyPhen, SIFT, PROVEAN, MutationAssesor, MutationTaster, LRT, GERP, SiPhy, phyloP and phastCons (34). Values larger than 0.5 are classified as "Likely Disease Causing."

MetaIR predicts the pathogenicity of missense mutations by integrating nine independent pathogenic scores and allele frequency using the logistic regression method. Variants are scored between 0 and 1, and the higher scored ones have a higher risk of being deleterious (35).

Mutation Assesor predicts the functional effect of amino acids by using evolutionary conservation of amino acids among homologous proteins. The program scores mutations between 0 and 1, then, classifies them as neutral, low, intermediate, or high. The higher the scores, the higher the risk of pathogenicity (36). Values greater than 0.93 are classified at high risk of pathogenicity.

I-Mutant is a neural network that calculates the protein stability change upon a single mutation. A protein sequence is given to the program in the FASTA format. I-Mutant can calculate the free-energy change ($\Delta\Delta G$) between mutant and wild-type proteins with the help of the FoldX program (37). According to the program output, if $\Delta\Delta G < 0$ then the mutation destabilizes the protein, and if $\Delta\Delta G > 0$ then the mutation stabilizes the protein.

PROVEAN predicts the effect of a single missense or indel mutation on the function of a protein. The program uses a fast algorithm, that allows calculations of all possible mutations of amino acids in mouse and human proteins (38). If the score is < 2.5, it destablizes the protein and its function.

FoldX is an algorithm that qualitatively predicts the interaction and stability of proteins upon mutation(s). The program enables a comparison between wild-type and mutant protein in the form of van der Waals interaction and reports a free energy change ($\Delta\Delta G$) (39). If $\Delta\Delta G$ reported by the program > 0, the mutation destabilizes the protein or interaction and if $\Delta\Delta G < 0$, the mutation stabilizes the protein or interaction.

FireDock is a web server that allows a refinement of protein-protein interaction by calculating binding energies via the molecular docking approach (40). The program performs local and global minimization by letting side chains find the optimum conformation and then calculates the binding energy of docked proteins. The more negative the binding energy, the more stable structure was.

Multiple sequence alignment was performed by using EMBL Multiple Sequence Comparison by the Log- Expectation (MUSCLE) server (41). Jalview software was used to visualize the sequence alignment results (42).

Open source Pymol software was used to visualize and generate protein figures (43).

RESULTS

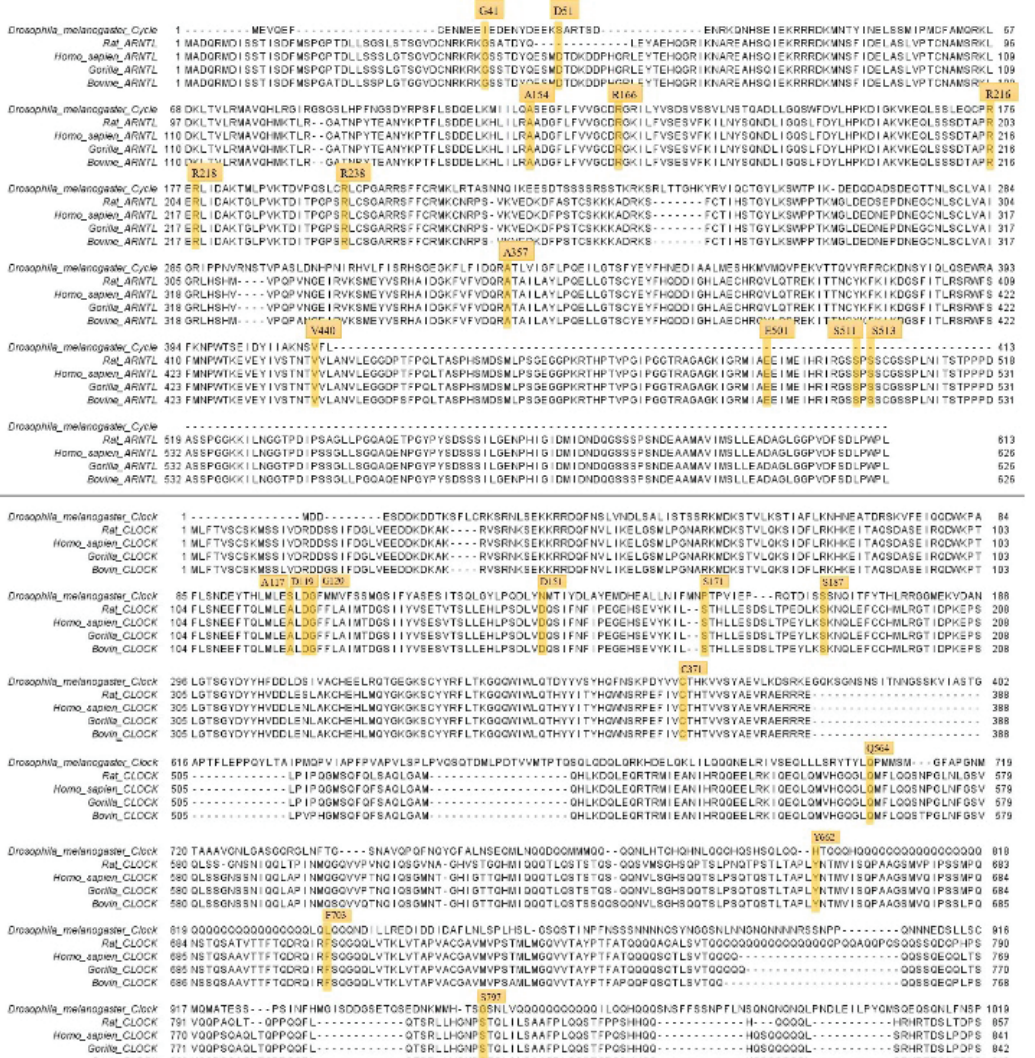
Initially, all CLOCK and BMAL1 SNPs were searched through the Ensembl database. For this purpose, we searched *"CLOCK"* and *"ARNTL"* (BMAL1) genes, one by one, in the "Human" species. Then, *CLOCK* (Human Gene) or *ARNTL* (Human Gene) were selected on the next page. "Variant table" on the left was selected and the "Filters" option appears. The following criteria were used to select SNPs: Class=SNP; Consequences=missense variant; SIFT=0-0.05; PolyPhen=0.9-1; CADD>30. More options can be found under the "Filter Other Columns" option. As a result, 71 variants for BMAL1 and 42 variants for CLOCK were obtained from the databank. Since BMAL1 is expressed in several isoforms, in the Ensembl database the same SNPs could be seen multiple times because the SNP was reported in different transcripts. We, therefore, selected ENST00000389707.8 transcripts (NP_001025443.1) for the sake of unifying the amino acid numbers in SNPs. A similar approach was used for CLOCK SNPs and ENST00000309964.8 transcript (NP_001254772.1) was selected. After eliminating the variants reported multiple times, we end up with 12 variants for BMAL1 and 12 variants for CLOCK. Multiple sequence alignment of BMAL1 and CLOCK among some vertebrates (e.g. bovin, gorilla, and rat) and *Drosophila melanogaster* which has a strong circadian clock system showed that almost all selected amino acids are conserved in these species (Figure 2). Since some in silico tools take advantage of available protein structures, we also classified variants as "structural" for those that are available on the crystal structures (PDB id: 4h10, 4f3l) and "non-structural" for those that were not resolved on the crystal. To determine "structural" variants we superimposed two crystal structures. Functional PAS and bHLH domains, and E-Box (DNA) were labeled on the aligned structures (Figures 1 and 3). Out of 12 BMAL1 variants, five of them were "structural" and seven of them were "non-structural". CLOCK had eight "structural," and four "non-structural" variants (Table 1).

The "Structural" SNPs of BMAL1 and CLOCK are shown on the BMAL1-CLOCK complex (Figure 3). Arg166 and Val440 are located on the surface of BMAL1 and are exposed to solvent. Ala154 and Arg238 are found on the BMAL1 PAS-A domain and both are interacting with CLOCK. Ala357 is found on the PAS-B domain and buried inside the BMAL1 (Figure 3A). Asp151, Ser171 and Ser187 are located on the surface of CLOCK and the PAS-A domain. Ala117, Asp 119 and Gly120 are also found on the PAS-A domain, however, located inside the CLOCK and interact with BMAL1. Cys371 is found in the CLOCK PAS-B domain and is located in a hydrophobic region (Figure 3B). The SNPs found on the BMAL1-CLOCK interaction site are shown with an asterisk (*) in Tables 2 and 3.

All SNPs SIFT, Poly-Phen, CADD, REVEL, MetalR, Mutation Assessor, and PhD-SNP values were evaluated. In addition, I-Mutant, PROVEAN, Dynamut, and FireDock software/web-servers were used for "structural" variants to evaluate the effect of mutations on BMAL1-CLOCK binding or change in the stability of proteins. Results are shown in Tables 2 and 3 for CLOCK and BMAL SNPs.

If the SIFT score is less than 0.05, it is classified as "Deleterious." If the Poly-Phen value is larger than 0.908, then it is evaluated as "Probably Damaging." If the CADD score is larger than 30, it is evaluated as "Deleterious". At the beginning of the SNP selection from the Ensembl database we used these three scores to select "Deleterious" mutations. All BMAL1 and CLOCK mutations were deleterious according to SIFT, Poly-Phen and CADD (Tables 2 and 3). Although there was no certain cut-off value for REVEL, scores larger than 0.5 are classified as "disease causing." Three structural BMAL1 SNPs (Arg166Gln, Val440Gly and Ala154Val) and six structural CLOCK SNPS (Cys371Gly, Ser171Phe, Gly120Val, Gly120Ser, Asp119Val and Ala117Val) were disease causing according to their REVEL values. The PhD-SNP probability score is calculated between 0 and 1; scores more than 0.5 are evaluated as pathogenic. One non-structural BMAL1 SNP (Asp51Tyr), and four structural BMAL1 SNPs (Arg166Gln, Val440Gly, Ala357Thr and Ala154Val) were evaluated as pathogenic by PhD-SNP (Table 2). All CLOCK SNPs, non-structural Ser797Leu and structural Asp151His, were classified as pathogenic by the same program (Table 3). I-Mutant calculated the free energy change ($\Delta\Delta G$) upon mutation in which mutants causing $\Delta\Delta G < 0$ are evaluated as destabilizing. All BMAL1 SNPs were determined as destabilizing by I-Mutant (Figure 4A). The same calculation for CLOCK SNPs showed that all except non-structural Tyr662Cys and structural Ser171Phe were destabilizing mutants (Table 3). PROVEAN is a useful tool to predict the effect of a single missense mutation on the stabilization and function of a protein. If the PROVEAN score is < 2.5, it can destabilize the protein and its function (38). All BMAL1 SNPs except two non-structural Glu501Lys and Gly41Arg SNPs turned out to be deleterious by the PROVEAN. The same analysis for CLOCK SNPs showed that all but one non-structural Ser797Leu and one structural Ser187Pro SNPs were deleterious. Finally, FireDock server was used to calculate the binding energy between wild-type BMAL1 and CLOCK, and between BMAL1-CLOCK variants found in the interaction site. Since FireDock calculates physical binding energy, lower binding energy means a stronger interaction. We calculated $\Delta\Delta G$ binding energy by subtracting the binding energy of wild-type BMAL1-CLOCK from the binding energy of SNPs ($\Delta\Delta G_{binding\ energy} = \Delta G_{SNP} - \Delta G_{WT}$). $\Delta\Delta G < 0$ means stabilized interaction between proteins. Interestingly, all SNPs found in the BMAL1-CLOCK interface stabilized the interaction. While both BMAL1 interface SNPs showed a strong effect (> 1 kcal/mol) on BMAL1-CLOCK binding energy, two out of four CLOCK interface SNPs showed a similar but less potent effect than BMAL1 SNPs on the binding energy (Figure 4B).

A



B

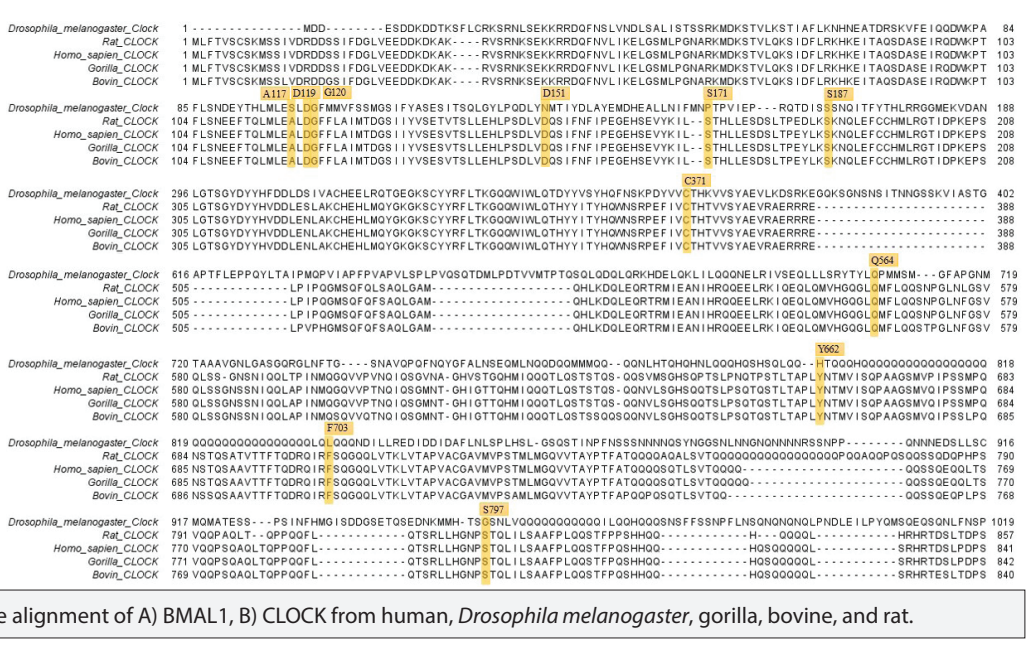


Figure 2. Multiple sequence alignment of A) BMAL1, B) CLOCK from human, *Drosophila melanogaster*, gorilla, bovine, and rat.

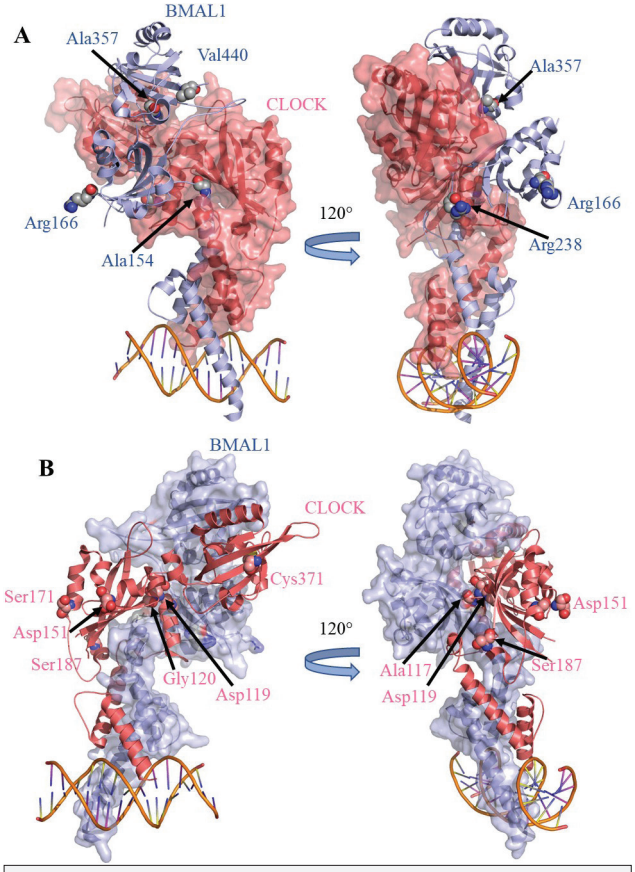


Figure 3. Position of "structural" SNPs of A) BMAL1 and B) CLOCK. In A) BMAL1 is shown in blue cartoon, SNPs are shown in sticks, CLOCK is shown in red surface representations. In B) CLOCK is shown in red cartoon, SNPs are shown in sticks, BMAL1 is shown in blue surface representations.

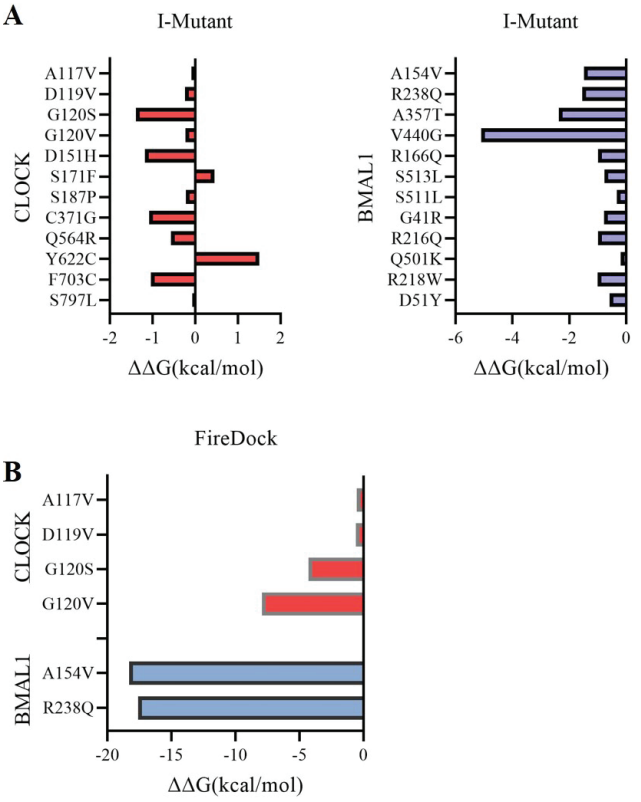


Figure 4. ΔΔG (kcal/mol) values calculated by A) I-Mutant for all SNPs, B) FireDock program for BMAL1-CLOCK interacting amino acids.

Table 1. List of all single nucleotide polymorphisms (SNPs) selected from the Ensembl database

	Structural		Non-structural	
	SNP ID	**Amino acid change**	**SNP ID**	**Amino acid change**
BMAL1	rs1226276621	Arg166Gln	rs1163932326	Asp51Tyr
	rs1450371323	Val440Gly	rs757010873	Arg218Trp
	rs746713814	Ala357Thr	rs146679326	Glu501Lys
	rs1434856001	Arg238GLn	rs368050146	Arg216Gln
	rs768337963	Ala154Val	rs771438108	Gly41Arg
			rs762595978	Ser511Leu
			rs1316220783	Ser513Leu
CLOCK	rs1406765260	Cys371Gly	rs953927923	Ser797Leu
	rs1204057040	Ser187Pro	rs1178386174	Phe703Cys
	rs768013881	Ser171Phe	rs763646149	Tyr662Cys
	rs375936156	Asp151His	rs752522729	Gln564Arg
	rs767225278	Gly120Val		
	rs1445208701	Gly120Ser		
	rs1289814329	Asp119Val		
	rs999814126	Ala117Val		

Table 2. Brain and muscle ARNT-Like 1 (BMAL1) single nucleotide polymorphisms (SNPs) and scores of in silico prediction tools.

	SNP ID	Amino acid change	SIFT	PolyPhen	CADD	REVEL	MetaLR	Mutation Assessor	PROVEAN	
Non-Structural	rs1163932326	Asp51Tyr	0	0.912	32	0.305	0.269	0.472	Disease	-3.02
	rs757010873	Arg218Trp	0	0.999	32	0.401	0.11	0.883	Neutral	-6.79
	rs146679326	Glu501Lys	0	0.991	32	0.25	0.24	0.819	Neutral	-2.1
	rs368050146	Arg216Gln	0	0.989	32	0.478	0.091	0.612	Neutral	-3.83
	rs771438108	Gly41Arg	0	0.999	30	0.379	0.179	0.535	Neutral	-2.11
	rs762595978	Ser511Leu	0.02	0.99	31	0.187	0.285	0.805	Neutral	-2.92
	rs1316220783	Ser513Leu	0.02	0.978	32	0.267	0.162	0.831	Neutral	-2.72
Structural	rs1226276621	Arg166Gln	0	0.996	31	0.521	0.26	0.745	Disease	-3.74
	rs1450371323	Val440Gly	0	0.971	30	0.702	0.16	0.695	Disease	-4.78
	rs746713814	Ala357Thr	0.01	0.998	33	0.371	0.14	0.800	Disease	-3.82
	rs1434856001*	Arg238GLn*	0.03	0.994	32	0.431	0.102	0.330	Neutral	-2.83
	rs768337963*	Ala154Val*	0.04	0.997	32	0.576	0.377	0.869	Disease	-3.11

Table 3. Circadian locomotor output cycles kaput (CLOCK) single nucleotide polymorphisms (SNPs) and scores of in silico prediction tools.

	SNP ID	Amino acid change	SIFT	PolyPhen	CADD	REVEL	MetaLR	Mutation Assessor	PROVEAN	
Non-Stuctural	rs953927923	Ser797Leu	0	0.966	32	0.339	0.069	0.480	Neutral	-1.83
	rs1178386174	Phe703Cys	0	0.962	31	0.495	0.062	0.700	Disease	-2.98
	rs763646149	Tyr662Cys	0.02	0.96	31	0.205	0.055	0.700	Disease	-3.31
	rs752522729	Gln564Arg	0	0.999	33	0.373	0.118	0.753	Disease	-2.703
Stuctural	rs1406765260	Cys371Gly	0	0.998	31	0.661	0.107	0.420	Disease	-10.831
	rs1204057040	Ser187Pro	0.02	0.932	32	0.142	0.043	0.385	Disease	-2.362
	rs768013881	Ser171Phe	0.01	0.969	31	0.639	0.163	0.906	Disease	-3.984
	rs375936156	Asp151His	0	0.976	31	0.431	0.088	0.237	Neutral	-4.39
	rs767225278*	Gly120Val*	0	0.999	32	0.705	0.249	0.950	Disease	-8.11
	rs1445208701*	Gly120Ser*	0	0.983	32	0.642	0.147	0.809	Disease	-5.34
	rs1289814329*	Asp119Val*	0	0.996	33	0.762	0.217	0.934	Disease	-8.24
	rs999814126*	Ala117Val*	0	0.99	32	0.678	0.107	0.898	Disease	-3.74

Table 4. Pathogenic score of all single nucleotide polymorphisms (SNPs). Pathogenic predictions are highlighted.

		SNP ID	Amino acid change	SIFT	PolyPhen	CADD	REVEL	Mutation Assessor	PhD-SNP	Provean	I-Mutant	Pathogenicity Score
BMAL1	Non-Stuctural	rs1163932326	Asp51Tyr	0	0.912	32	0.305	0.472	Disease	Deleterious	-0.59	6
		rs757010873	Arg218Trp	0	0.999	32	0.401	0.883	Neutral	Deleterious	-1.00	5
		rs146679326	Glu501Lys	0	0.991	32	0.25	0.819	Neutral	Neutral	-0.20	4
		rs368050146	Arg216Gln	0	0.989	32	0.478	0.612	Neutral	Deleterious	-0.99	5
		rs771438108	Gly41Arg	0	0.999	30	0.379	0.535	Neutral	Neutral	-0.78	4
		rs762595978	Ser511Leu	0.02	0.99	31	0.187	0.805	Neutral	Deleterious	-0.33	5
		rs1316220783	Ser513Leu	0.02	0.978	32	0.267	0.831	Neutral	Deleterious	-0.76	5
	Stuctural	rs1226276621	Arg166Gln	0	0.996	31	0.521	0.745	Disease	Deleterious	-0.99	**7**
		rs1450371323	Val440Gly	0	0.971	30	0.702	0.695	Disease	Deleterious	-5.10	**7**
		rs746713814	Ala357Thr	0.01	0.998	33	0.371	0.800	Disease	Deleterious	-2.37	6
		rs1434856001*	Arg238GLn*	0.03	0.994	32	0.431	0.330	Neutral	Deleterious	-1.54	5
		rs768337963*	Ala154Val*	0.04	0.997	32	0.576	0.869	Disease	Deleterious	-1.48	**7**
CLOCK	Non-Structural	rs953927923	Ser797Leu	0	0.966	32	0.339	0.480	Neutral	Neutral	-0.07	4
		rs1178386174	Phe703Cys	0	0.962	31	0.495	0.700	Disease	Deleterious	-1.04	6
		rs763646149	Tyr662Cys	0.02	0.96	31	0.205	0.700	Disease	Deleterious	1.50	5
		rs752522729	Gln564Arg	0	0.999	33	0.373	0.753	Disease	Deleterious	-0.57	6
	Stuctural	rs1406765260	Cys371Gly	0	0.998	31	0.661	0.420	Disease	Deleterious	-1.08	**7**
		rs1204057040	Ser187Pro	0.02	0.932	32	0.142	0.385	Disease	Neutral	-0.22	5
		rs768013881	Ser171Phe	0.01	0.969	31	0.639	0.906	Disease	Deleterious	0.45	6
		rs375936156	Asp151His	0	0.976	31	0.431	0.237	Neutral	Deleterious	-1.18	5
		rs767225278*	Gly120Val*	0	0.999	32	0.705	0.950	Disease	Deleterious	-0.23	**8**
		rs1445208701*	Gly120Ser*	0	0.983	32	0.642	0.809	Disease	Deleterious	-1.39	**7**
		rs1289814329*	Asp119Val*	0	0.996	33	0.762	0.934	Disease	Deleterious	-0.24	**8**
		rs999814126*	Ala117Val*	0	0.99	32	0.678	0.898	Disease	Deleterious	-0.09	**7**

DISCUSSION

The SNP is a single nucleotide difference among gene sequences of the canonical population. SNPs causing missense mutation to change an amino acid to a different one. Some SNPs and exon skipping mutations of core circadian clock genes are related to sleep and mood disorders (24-26,44,45). Specifically, a PER2 SNP (Ser662Gly) was detected in the population having ASPS (24). Similarly, a CRY2 SNP causing Ala260Thr mutation was also associated with ASPA (45). In common, both PER2 and CRY2 mutations destabilized the protein and changed the timing of the circadian clock. In addition to understanding the disease mechanism, characterization of rare SNPs helps to understand the clock mechanism better. A recent study, analyzing the effect of CRY1 SNPs on the clock mechanism, showed that CRY1 is an allosterically regulated protein (27). Another study dealing with the effect of CRY2 SNPs showed that secondary pocket amino acids are important for nuclear localization and stability of CRY2 (28). Thus, understanding the effect of clock genes' SNPs on the circadian clock is an active research field.

For this purpose, we identified pathogenic missense SNPs in the Ensembl database by using various in silico tools. Our initial analysis by using pathogenicity criteria of SIFT, PolyPhen, and CADD showed that there were 12 SNPs for both BMAL1 and CLOCK in the Ensembl database (Table 1). Since some software used protein structures for analysis, we classified SNPs as "structural" and "non-structural" according to the presence of amino acids available on crystal structures (pdb id: 4f3l, 4h10) (Table 1). As a result, seven non-structural and five structural SNPs for

BMAL1; and four structural and eight non-structural SNPs for CLOCK were determined (Table1) (Figure 3). Multiple sequence alignment revealed that in vertebrates all analyzed amino acids of BMAL1 were conserved except Asp51 in rat (Figure 2A). In drosophila, Gly41 and Asp51 are replaced with isoleucine and serine, respectively. In addition to these two amino acids, *Cycle* protein has a lack of C-terminal and only 413 amino acids; Glu501, Ser511, and Ser513 of human BMAL1 do not exist in the fly *Cycle* protein. Among CLOCK proteins analyzed in Figure 2B, all studied SNP amino acids are conserved in vertebrates. However, only Asp119, Gly120, Ser187, Cys371, and Gln564 amino acids are conserved in the *Clock* protein of *Drosophila melanogaster* (Figure 2B). Other than Ala117 of fly *Clock*, all SNPs found on the BMAL1-CLOCK interaction interface are conserved among the analyzed species (Figure 2).

The REVEL program predicted only three structural Arg166Gln, Val440Gly and Ala154Val BMAL1 SNPs, and six structural Cys371Gly, Ser171Phe, Gly120Val, Gly120Ser, Asp119Val and Ala117Val CLOCK SNPs as disease-causing. The PhD-SNP software predicted that one non-structural Asp51Tyr, and four structural Arg166Gln, Val440Gly, Ala357Thr and Ala154Val BMAL1 SNPs as pathogenic (Table 2). The same program classified all CLOCK SNPs other than non-structural Ser797Leu and structural Asp151His as pathogenic (Table 3). The MetalR program uses multiple in silico prediction programs and scores the mutation. Although there is no certain cut-off score to evaluate mutations as pathogenic by the MetalR program, a larger score means a higher risk of being pathogenic. PROVEAN predicts that all BMAL1 SNPs except two non-structural Glu501Lys and Gly41Arg were deleterious. The same program predicted all but one non-structural Ser797Leu and one structural Ser187Pro SNPs of CLOCK as deleterious. The mutation assessor classified only two structural CLOCK SNPs Asp119Val and Gly120Val as pathogenic.

According to I-Mutant, if the free energy change (ΔΔG) upon mutation < 0, it is a destabilizing mutation. The I-Mutant predicted that all BMAL1 SNPs were destabilizing (Figure 4A). Similarly, all CLOCK SNPs, other than Ser171Phe and Tyr622Cys were destabilizing mutants (Figure 4A). To better understand the I-Mutant results we analyzed proteins according to the hydrophobicity of amino acids. Ala357 and Val440 of BMAL1 are buried in a hydrophobic region. Since glycine and threonine are less hydrophobic than valine and alanine, respectively, disruption of these hydrophobic cores might destabilize the protein. Ala154 is remarkably close to a region rich in basic amino acids. Thus, the Ala154Val mutation that introduced a more hydrophobic amino acid may destabilize the protein. The side chains of Ala117 and Gly120 of CLOCK are located near a hydrophobic region. Therefore, Ala117Val and Gly120Val mutations which are also hydrophobic amino acids did not destabilize the protein. However, Gly120Ser mutation which introduced a polar amino acid destabilized the protein. When Asp151 is exposed to solvent and because of its negative charge, it can be found in H-bond interactions with surrounding water molecules. Asp151His mutation may decrease the number and strength of H-bond interactions since histidine is mostly neutral around the physiological pH. Therefore, Asp151His mutation is a destabilizing mutation.

To predict the binding energy change in BMAL1-CLOCK interaction upon introducing SNPs, we used the FireDock server. The server calculates physical binding energy and if ΔΔG after mutation < 0, it means a stabilized energy (in contrast to I-Mutant). FireDock predicted that all SNPs found on the interaction interface were stabilizing the BMAL1-CLOCK interaction (Figure 4B). It showed that a gain of function mutation on BMAL1 and CLOCK caused a stronger interaction between them, but it did not affect the circadian rhythm in cell-based circadian assays. Similarly, the same mutations had a subtle effect on BMAL1-CLOCK dependent transactivation (29). On the other hand, loss of function mutations that weakened the BMAL1-CLOCK binding significantly abolished the transactivation (29). Since the FireDock analysis predicted that all SNPs had a better binding energy than the wild-type, these mutants should not have a detrimental effect on the BMAL1-CLOCK transactivation.

Finally, we generated a table that collectively gives a score for each analysis. If an SNP was predicted as pathogenic or destabilizing by a program, the SNP gets a score of 1. A higher overall score means a higher risk of pathogenicity. For this, we used results from the SIFT, PolyPhen, CADD, REVEL, PhD-SNP, PROVEAN, Mutation Assessor and I-Mutant analysis. So, an SNP could get the highest score of 8. Our results showed that structural CLOCK SNPs: Gly120Val and Asp119Val got the highest score of 8; Cys371Gly, Gly120Ser, and Ala117Val got 7; BMAL1 structural SNPs Arg166Gln, Val440Gly, and Ala154Val also got 7 (Table 4). The highest scored SNPs can have a significant effect on the BMAL1-CLOCK activity and may disrupt the robust ticking of the circadian clock mechanism.

Despite the fact that Gly41Arg SNP got a pathogenicity score of 4, the amino acid is at the end of the nuclear localization signal sequence which is important for nuclear translocation of BMAL1 and CLOCK (46). Since BMAL1 carries CLOCK to the nucleus, if BMAL1 cannot translocate to the nucleus neither can CLOCK. So, if this SNP interferes with BMAL1 translocation, it may disrupt the clock mechanism.

Cys371Gly obtained a pathogenicity score of 7 and should be analyzed in detail. Recently, it was found that the side chain of Cys195 (-SH) of CLOCK can be oxidized to sulfenic acid (-SOH) depending on the amount of H_2O_2 (47). As a result, oxidized Cys195 stabilized the CLOCK-BMAL1 interaction and enhanced the transactivation. Since all cysteine amino acids can be oxidized, this possibility should be considered while experimentally analyzing the effect of Cys371Gly.

Heme can bind to CLOCK and may interfere with CLOCK binding to E-Box and regulate the circadian clock (48,49). His141 was critical for Heme binding. In the CLOCK amino acids we studied, Ala117, Asp119, Gly120, and Asp151 surrounded His141 in the crystal structure. These SNPs can affect the orientation of the His141 side chain and heme binding.

CONCLUSION

Among the thousands of SNPs in the Ensemble database, we filtered out only missense mutation with high pathogenic probability. After applying pathogenic criteria of SIFT, PolyPhen, and CADD, we studied 12 SNPs using various in silico tools. In spite of the fact that these in silico tool analysis were not adequate to make general conclusions about the effect of SNPs on biological systems, they are useful to narrow down the number of SNPs for experimental studies. Thus, the in silico analysis help to save time and budget. Researchers aiming to experimentally analyze the effect of CLOCK and BMAL1 SNPs on the clock mechanism should give priority to those highlighted in this study.

Peer Review: Externally peer-reviewed.

Conflict of Interest: Author declared no conflict of interest.

Financial Disclosure: Author declared no financial support.

REFERENCES

1. Dunlap JC. Molecular bases for circadian clocks. Cell 1999; 96(2): 271-90.
2. Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, et al. Coordinated transcription of key pathways in the mouse by the circadian clock. Cell 2002; 109(3): 307-20.
3. Sancar G, Brunner M. Circadian clocks and energy metabolism. Cell Mol Life Sci 2014; 71(14): 2667-80.
4. Turek FW, Joshu C, Kohsaka A, Lin E, Ivanova G, McDearmon E, et al. Obesity and metabolic syndrome in circadian Clock mutant mice. Science 2005; 308(5724): 1043-45.
5. Nernpermpisooth N, Qiu SQ, Mintz JD, Suvitayavat W, Thirawarapan S, Rudic DR, et al. Obesity Alters the Peripheral Circadian Clock in the Aorta and Microcirculation. Microcirculation 2015; 22(4): 257-66.
6. Damiola F, Le Minh N, Preitner N, Kornmann B, Fleury-Olela F, Schibler U. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. Gene Dev 2000; 14(23): 2950-61.
7. Mohawk JA, Takahashi JS. Cell autonomy and synchrony of suprachiasmatic nucleus circadian oscillators. Trends Neurosci 2011; 34(7): 349-58.
8. Kavakli IH, Sancar A. Circadian photoreception in humans and mice. Mol Interv 2002; 2(8): 484-92.
9. Ko CH, Takahashi JS. Molecular components of the mammalian circadian clock. Hum Mol Genet 2006; 15: R271-R77.
10. Welsh DK, Logothetis DE, Meister M, Reppert SM. Individual Neurons Dissociated from Rat Suprachiasmatic Nucleus Express Independently Phased Circadian Firing Rhythms. Neuron 1995; 14(4): 697-706.
11. King DP, Zhao YL, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP, et al. Positional cloning of the mouse circadian Clock gene. Cell 1997; 89(4): 641-53.
12. Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, Mcdonald JD, et al. Mutagenesis and Mapping of a Mouse Gene Clock, Essential for Circadian Behavior. Science 1994; 264 (5159): 719-25.
13. Bunger MK, Wilsbacher LD, Moran SM, Clendenin C, Radcliffe LA, Hogenesch JB, et al. Mop3 is an essential component of the master circadian pacemaker in mammals. Cell 2000; 103(7): 1009-17.
14. Shearman LP, Zylka MJ, Weaver DR, Kolakowski LF, Reppert SM. Two period homologs: Circadian expression and photic regulation in the suprachiasmatic nuclei. Neuron 1997; 19(6): 1261-69.
15. Griffin EA, Staknis D, Weitz CJ. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. Science 1999; 286(5440): 768-71.
16. Kume K, Zylka MJ, Sriram S, Shearman LP, Weaver DR, Jin XW, et al. mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. Cell 1999; 98(2): 193-205.
17. Hao HP, Allen DL, Hardin PE. A circadian enhancer mediates PER-dependent mRNA cycling in Drosophila melanogaster. Mol Cell Biol 1997; 17(7): 3687-93.
18. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: Implications for biology and medicine. P Natl Acad Sci USA 2014; 111(45): 16219-24.
19. Hirano A, Fu YH, Ptacek LJ. The intricate dance of post-translational modifications in the rhythm of life. Nat Struct Mol Biol 2016; 23(12): 1053-60.
20. Preitner N, Damiola F, Molina LL, Zakany J, Duboule D, Albrecht U, et al. The orphan nuclear receptor REV-ERB alpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. Cell 2002; 110(2): 251-60.
21. Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, McNamara P, et al. A functional genomics strategy reveals rora as a component of the mammalian circadian clock. Neuron 2004; 43(4): 527-37.
22. Kelleher FC, Rao A, Maguire A. Circadian molecular clocks and cancer. Cancer Lett 2014; 342(1): 9-18.
23. Scheer FAJL, Hilton MF, Mantzoros CS, Shea SA. Adverse metabolic and cardiovascular consequences of circadian misalignment. P Natl Acad Sci USA 2009; 106(11): 4453-58.
24. Fu Y, Jones CR, Toh K, Virshup D, Ptacek LJ. An hPer2 phosphorylation site mutation in familial Advanced Sleep-Phase Syndrome. Am J Hum Genet 2001; 69(4): 597-97.
25. Patke A, Murphy PJ, Onat OE, Krieger AC, Ozcelik T, Campbell SS, et al. Mutation of the Human Circadian Clock Gene CRY1 in Familial Delayed Sleep Phase Disorder. Cell 2017; 169(2): 203-15.
26. Onat OE, Kars ME, Gul S, Bilguvar K, Wu YM, Ozhan A, et al. Human CRY1 variants associate with attention deficit/hyperactivity disorder. J Clin Invest 2020; 130(7): 3885-900.
27. Gul S, Aydin C, Ozcan O, Gurkan B, Surme S, Baris I, et al. The Arg-293 of Cryptochrome1 is responsible for the allosteric regulation of CLOCK-CRY1 binding in circadian rhythm. J Biol Chem 2020; 295(50): 17187-99.
28. Parlak GC, Camur BB, Gul S, Ozcan O, Baris I, Kavakli IH. The secondary pocket of Cryptochrome 2 is important for the regulation of its stability and localization. J Biol Chem 2022: 102334.
29. Huang NA, Chelliah Y, Shan YL, Taylor CA, Yoo SH, Partch C, et al. Crystal Structure of the Heterodimeric CLOCK:BMAL1 Transcriptional Activator Complex. Science 2012; 337(6091): 189-94.
30. Wang ZX, Wu YL, Li LF, Su XD. Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA. Cell Res 2013; 23(2): 213-24.
31. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003; 31(13): 3812-14.
32. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002; 30(17): 3894-900.
33. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014; 46(3): 310.
34. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet 2016; 99(4): 877-85.

35. Dong CL, Wei P, Jian XQ, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 2015; 24(8): 2125-37.
36. Li C, Zhi D, Wang K, Liu X. MetaRNN: Differentiating Rare Pathogenic and Rare Benign Missense SNVs and InDels Using Deep Learning. bioRxiv 2021.
37. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005; 33: W306-W10.
38. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015; 31(16): 2745-47.
39. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res 2005; 33: W382-W88.
40. Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ. FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res 2008; 36: W229-W32.
41. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res 2022; 50(W1): W276-W79.
42. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009; 25(9): 1189-91.
43. Schrodinger L. The PyMOL Molecular Graphics System. 2010.
44. Hua P, Liu WG, Chen DH, Zhao YY, Chen L, Zhang N, et al. Cry1 and Tef gene polymorphisms are associated with major depressive disorder in the Chinese population. J Affect Disorders 2014; 157: 100-03.
45. Hirano A, Shi G, Jones CR, Lipzen A, Pennacchio LA, Xu Y, et al. A Cryptochrome 2 mutation yields advanced sleep phase in humans. Elife 2016; 5.
46. Kwon I, Lee J, Chang SH, Jung NC, Lee BJ, Son GH, et al. BMAL1 shuttling controls transactivation and degradation of the CLOCK/BMAL1 heterodimer. Mol Cell Biol 2006; 26(19): 7318-30.
47. Pei JF, Li XK, Li WQ, Gao Q, Zhang Y, Wang XM, et al. Diurnal oscillations of endogenous H2O2 sustained by p66(Shc) regulate circadian clocks. Nat Cell Biol 2019; 21(12): 1553-64.
48. Shimizu T, Huang D, Yan F, Stranava M, Bartosova M, Fojtikova V, et al. Gaseous O2, NO, and CO in signal transduction: structure and function relationships of heme-based gas sensors and heme-redox sensors. Chem Rev 2015; 115(13): 6491-533.
49. Freeman SL, Kwon H, Portolano N, Parkin G, Venkatraman Girija U, Basran J, et al. Heme binding to human CLOCK affects interactions with the E-box. Proc Natl Acad Sci U S A 2019; 116(40): 19911-16.