Progress in Nutrition 2022; Vol. 24, N.1: e2022002 DOI 10.23751/pn.v24i1.11487

Original article

Determination of the Structural and Functional Impact of High-Risk Missense SNPs in the Obesity-Associated Gene, Beta-Catenin-Like Protein 1 (*CTNNBL1*), by Bioinformatic Methods

Orcun Avsar
Hitit University, Department of Molecular Biology and Genetics, Corum, Turkey

Summary. Obesity with a complex etiology is significantly correlated with mortality and morbidity and it is becoming an epidemic through the worldwide. Moreover, obesity is a great risk factor for various diseases such as cancer, diabetes, and chronic diseases. CTNNBL1 protein belongs to armadillocontaining protein family and is a member of the spliceosome of pre-mRNA processing factor 19 (Prp19) through the interaction with N-terminal sequence of CDC5L. It has been reported that beta-catenin-like protein 1 is associated with body fat mass and Body Mass Index (BMI). The underlying molecular mechanisms of obesity is needed to be clarified by comprehensive studies. Therefore, we aimed to conduct an *in silico* study to identify the effects of the high-risk nsSNPs of obesity-associated *CTNNBL1* gene. Deleterious missense SNPs were analyzed through five different bioinformatic tools. Only eight missense SNPs were found to be deleterious and further investigated for prediction their effects on protein stability, the evolutionary dynamics of the changes of amino acids, PTMs sites, and structure conformation. Based on the web-based bioinformatic tools, we demonstrated that only five nsSNPs R323W (rs141919968), R73Q (rs144576870), R458H (rs151227978), P274H (rs200219582), and N299S (rs369990971) from eight *CTNNBL1* deleterious nsSNPs were significant. No study conducted with these deleterious nsSNPs is reported in the literature. We suppose that the predicted pathogenic nsSNPs may be implicated in the pathogenesis of obesity and confirmation of the results by population studies are required.

Key words: Obesity, CTNNBL1, Protein stability, SNP, Mutation

Introduction

Obesity is a chronic and multifactorial disease correlated with excessive body fat accumulation (1). Obesity is a serious public health problem and associated with various diseases such as cancer and cardiovascular diseases. Various factors such as genetic background, epigenetic modifications, diet, environmental factors, and exercise are implicated in the pathogenesis of obesity (2, 3). All social class and ages are increasingly affected by obesity in the world (1).

Beta-catenin-like protein 1 (CTNNBL1) gene is localized on the human chromosome 20q11.23 and involves 16 exons and 15 introns. Beta-catenin-like protein 1 is a nuclear protein with a highly conserved properties among mammalians (4). CTNNBL1 protein belongs to armadillocontaining protein family and is a member of the spliceosome of pre-mRNA processing factor 19 (Prp19) through the interaction with N-terminal sequence of CDC5L (5). Beta-catenin-like protein 1 is a component of spliceosomal complex and implicated in several biological processes such as

antibody diversification, apoptosis, and regulation of cell-cell adhesion (4). *CTNNBL1* gene is expressed in all tissues of the human body, but especially higly expressed in skeletal muscle (significant for energy metabolism), thyroid, heart, testes, placenta, and spleen (6).

Beta-catenin-like protein 1 shares similar biological functions with beta-catenin that is a part of the pathway of Wnt-signaling. The Wnt-signaling is significantly involved in cellular proliferation, development, and differentiation in various organisms. The pathway of Wnt/b-catenin-signaling may inhibit the gene expression associated with pre-adipocyte differentiation (6, 7, 8). It has been reported that beta-catenin-like protein 1 is associated with body fat mass and Body Mass Index (BMI) (6). Genetic variants in human *CTNNBL1* gene was suggested to be associated with elevated body fat mass and obesity (9). In a study conducted with Danish people, it was determined that *CTNNBL1* gene variants were related to height and body weight (10).

Obesity is increasingly affecting individuals in the worldwide and the genetic determinants are needed to be comprehensively elucidated in obesity. In the present study, we aimed to investigate the pathogenic effects of missense variants in the *CTNNBL1* gene and to determine the risk alleles for obesity.

Material and methods

The missense SNPs and associated protein sequence for the human *CTNNBL1* gene were retrieved from NCBI dbSNP database. Then, the missense SNPs were subjected to numerous *in silico* analyses.

Analysis of functional non-synonymous SNPs

To predict and analyze the pathogenicity of human *CTNNBL1* nsSNPs, five bioinformatic programs were used involving PANTHER-PSEP, PROVEAN, PhD-SNP, SIFT, SNAP2.

PANTHER-PSEP predicts missense genetic variants which can have causative effect on human diseases. This freely available web tool estimates potential deleterious or damaging nsSNPs according to evolutionary conservation scores (11). PROVEAN algorithmic program predicts the impacts of single or multiple amino acid changes, deletions, and insertions. It enables the rapid analysis for mouse and human variants at protein and genetic levels (12). PhD-SNP determines single point mutations as "disease" or "neutral" based on support vector machines (SVMs) (13). The Sorting Intolerant from Tolerant (SIFT) bioinformatic tool estimates the functional effects of genetic variants in coding sequences on protein. This web server that characterizes missense mutations is with the cutoff score 0.05 and the variations with scores > 0.05 are estimated to be pathogenic (14). SNAP2 web server categorizes missense mutations as "effective" and "neutral". SNAP2 scores are in the range of -100 (significantly neutral) and +100 (significantly effective) (15).

Prediction of beta-catenin-like protein 1 stability

Beta-catenin-like protein 1 was subjected to stability analysis by the use of I-Mutant 3.0 and MUpro. I-Mutant 3.0 estimates the effects of single nucleotide polymorphisms on protein stability based on SVMs. Stability prediction is performed based on protein primary structure and associated with free energy change value ($\Delta\Delta G$). The more negative $\Delta\Delta G$ value is correlated with more decreased protein stability in mutant model (16). MUpro that is based on Neural Networks and Support Vector Machines predicts how mutation affects protein stability with the confidence score in the range of -1 and +1. The calculated score > 0 demonstrates increased protein stability, whereas the score < 0 shows the decreased stability (17).

Evolutionary conservation prediction

The ConSurf web server was used to evaluate the evolutionary rates of amino acids in human beta-catenin-like protein 1. The widely used prediction tool determines the evolutionary dynamics of the changes of amino acids according to the phylogenetic associations. ConSurf calculates the conservation scores of each amino acid. The amino acid residues with the conservation scores in the range of 7 and 9 indicate evolutionary conservative ones (18).

Post-translational modification prediction by MusiteDeep

MusiteDeep was used to determine the post-translational modification (PTM) sites in human beta-catenin-like protein 1. The online tool provides the visualization and prediction of PTM regions in protein based on protein sequence (19).

Amino acid substitution modeling by Project HOPE software

The HOPE software determines the impacts of point mutations on protein function and conformation of structure. This software enables to predict the effects of wild-type and mutant amino acids on 3D-structure of protein through modeling (20).

Results

Total 39935 SNPs in the human *CTNNBL1* gene were reported in the NCBI dbSNP database. Among these SNPs, 165 were synonymous SNPs and 325 were missense SNPs. Pathogenic nsSNPs may affect the human beta-catenin-like protein 1 structurally and functionally therefore, the 325 missense SNPs were exported for our further investigation.

Five different algorithmic programs were used for the structural and functional evaluation of missense SNPs. While one bioinformatic tool calls some missense SNPs as neutral, another program may determine them as deleterious. The results retrieved from these bioinformatic tools are seen in Table 1. 8 missense SNPs out of 325 which were predicted as deleterious by at least 4 algorithmic tools are: rs141919968 (R323W), rs144576870 (R73Q), rs151227978 (R458H), rs200219582 (P274H), rs200874158 (K365N), rs369990971 (N299S), rs370193565 (N299S), and rs370578340 (R202W).

These 8 nsSNPs were further investigated for the prediction of the effects on protein stability through I-Mutant 3.0 and MUpro and the results are shown in Table 2. It was determined that these missense SNPs decrease or increase the stability of human beta-catenin-like protein 1. Based on the findings of the servers, nsSNPs rs141919968, rs144576870, rs151227978, rs200219582, rs200874158, rs369990971, rs370193565, and rs370578340 had a DDG value < -0.05 and it was demonstrated that these nsSNPs are significantly unstable and lead to decreased beta-catenin-like protein 1 stability.

The amino acids that are located at particular positions in protein sequence and vital for the accurate function of a protein alter more slowly compared to the other residues and these amino acids are called as "evolutionarily conserved". Moreover, it is known that damaging mutations occur more frequently in conserved regions of a protein (21). The 8 nsSNPs analyzed by ConSurf predicted that R323W, R73Q, R458H, P274H, and N299S were highly conserved and these residues are determined as functional residues. The results of ConSurf web server is seen in Figure 1.

Post-translational modifications (methylation, acetylation, glycosylation, phosphorylation, SUMOylation, ubiquitination, hydroxylation, palmitoylation, pyrrolidone carboxylic acid)mediated by missense SNPs were identified using MusiteDeep web tool. According to MusiteDeep, no PTMs were predicted at residues associated with deleterious mutations, and the findings are shown in Table 3, Figure 2, and Figure 3.

Amino acids have special physicochemical characteristics such as charge, molecular weight, and hydrophobicity value. It was shown that some of these properties differed between mutant and wild-type amino acids by Project HOPE software and these findings are given in Table 4. Furthermore, 3D-modeling of CTNNBL1 based on the selected high-risk deleterious nsSNPs are shown in Table 5.

Discussion

Obesity with a complex etiology is significantly correlated with mortality and morbidity and it is becoming an epidemic through the worldwide. Furthermore, obesity is a great risk factor for various diseases such as cancer, diabetes, and chronic diseases. Therefore, obesity management is crucial for public health (22). In this regard, the underlying molecular mechanisms of obesity is needed to be clarified by comprehensive

Table 1. nsSNPs of human *CTNNBL1* gene predicted as pathogenic by various bioinformatics methods

SNP ID	Nucleotide substitution	Amino acid change	PANTHER-PSEP	PANTHER-PSEP Pdel	PROVEAN result	PROVEAN score	PhD-SNP prediction	PhD-SNP RI	SIFT result	SIFT score	SNAP2 Prediction	SNAP score
rs141919968	C/T	R323W	Pro-damg	0.89	Disease	-6.840	Disease	3	DEL	0.002	Effect	89
rs144576870	G/A	R73Q	Pro-damg	0.89	Disease	-3.626	Disease	1	DEL	0	Effect	77
rs151227978	G/A	R458H	Pro-damg	0.89	Disease	-4.470	Neutral	1	DEL	0.003	Effect	13
rs200219582	C/A	P274H	Pro-damg	0.89	Disease	-8.700	Neutral	2	DEL	0	Effect	63
rs200874158	G/T	K365N	Pro-damg	0.85	Disease	-3.153	Disease	1	DEL	0.008	Effect	24
rs369990971	A/G	N299S	Pro-damg	0.89	Disease	-4.875	Neutral	4	DEL	0	Effect	41
rs370193565	C/T	R517W	Pro-damg	0.85	Neutral	-2.331	Disease	2	DEL	0.016	Effect	12
rs370578340	C/T	R202W	Pro-damg	0.85	Disease	-5.966	Disease	2	DEL	0.001	Effect	73

Table 2. The results of protein stability prediction by I-Mutant and MUpro

SNP ID	Amino acid change	I-Mutant prediction	I-Mutant RI	MUpro prediction	Delta Delta G (DDG)
rs141919968	R323W	Decrease	3	Decrease	-0.6614823
rs144576870	R73Q	Decrease	9	Decrease	-0.84201248
rs151227978	R458H	Decrease	9	Decrease	-1,26
rs200219582	P274H	Decrease	6	Decrease	-1,27
rs200874158	K365N	Decrease	1	Decrease	-1,06
rs369990971	N299S	Decrease	3	Decrease	-1,41
rs370193565	R517W	Decrease	4	Decrease	-0.57566807
rs370578340	R202W	Decrease	4	Decrease	-0.73730608

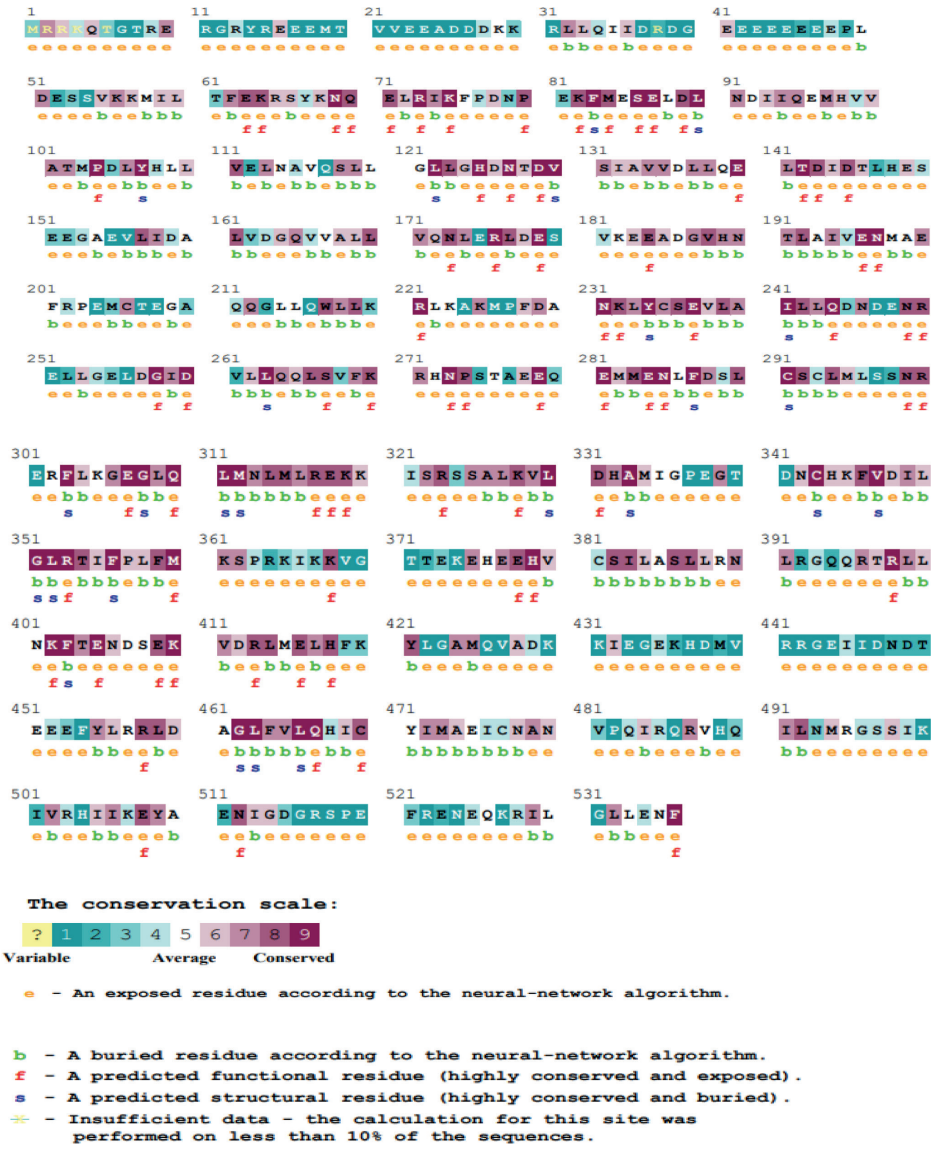


Figure 1. Evolutionary conservancy of human CTNNBL1 created by ConSurf server

Table 3. Prediction of post-translational modifications at residues caused by high-risk nsSNPs

SNP ID	Amino acid change	PTM scores (Cutoff=0.5)	Prediction
rs141919968	R323W	Methylarginine:0.034	None
rs144576870	R73Q	Methylarginine:0.024	None
rs151227978	R458H	Methylarginine:0.016	None
rs200219582	P274H	Hydroxyproline:0.05	None
rs200874158	K365N	Ubiquitination:0.18; SUMOylation:0.046; N6-acetyllysine:0.165; Methyllysine:0.057; Hydroxylysine:0.046	None
rs369990971	N299S	N-linked glycosylation:0.036	None
rs370193565	R517W	Methylarginine:0.039	None
rs370578340	R202W	Methylarginine:0.022	None

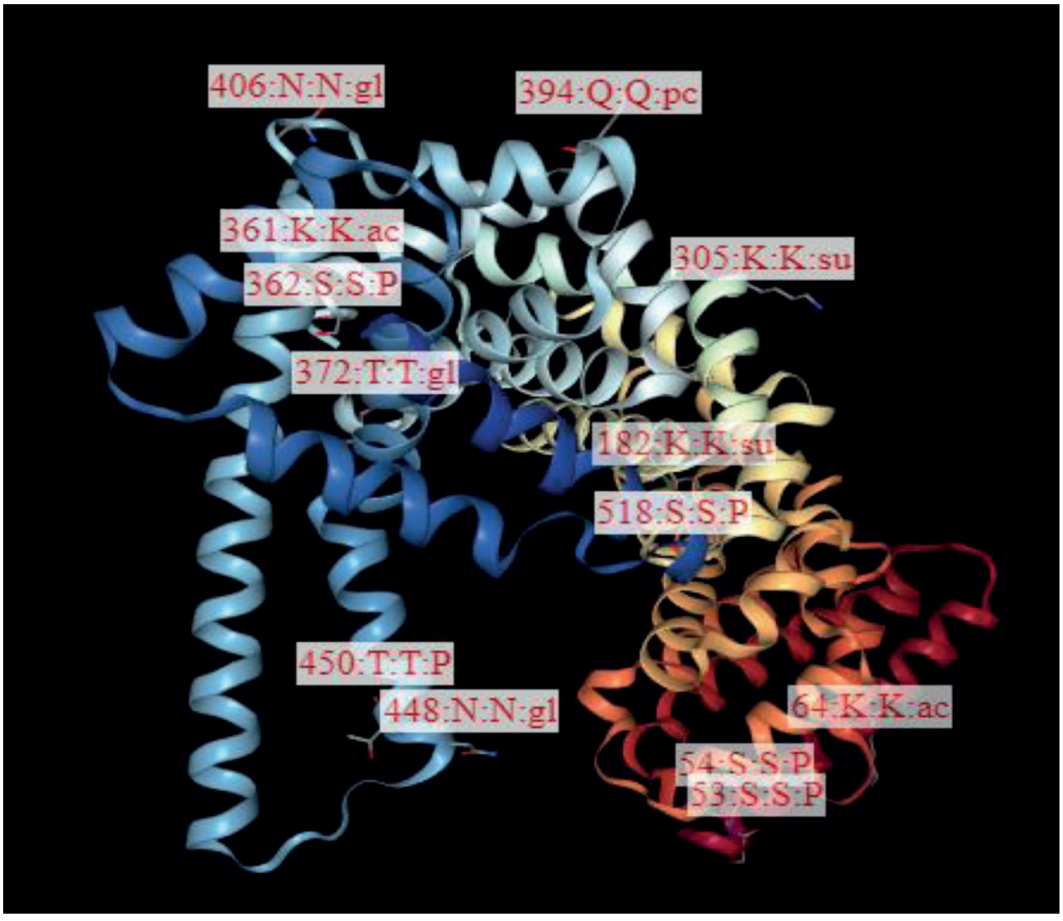


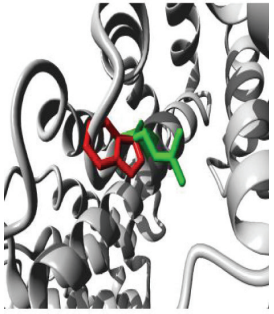
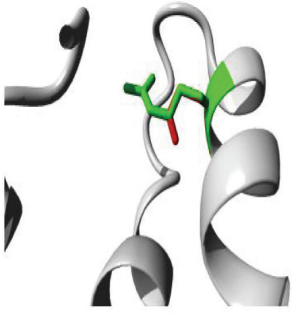
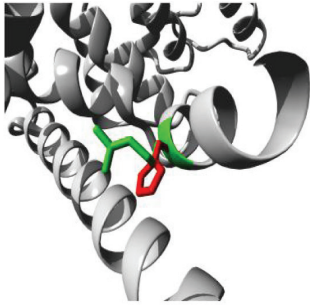
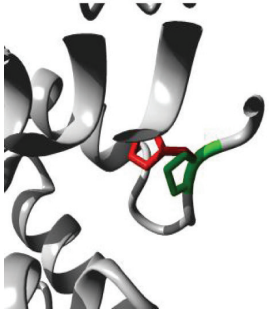
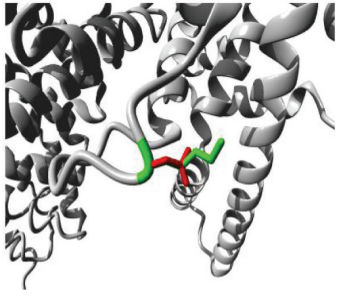
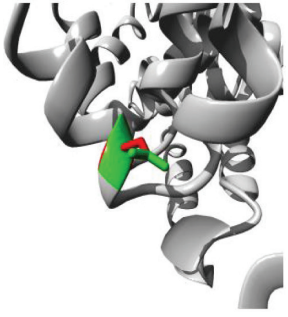
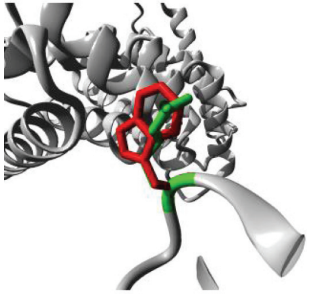
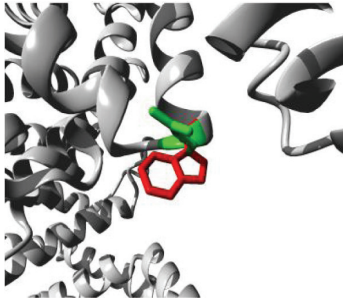
Figure 2. Predicted post-translational modification sites in 3D structure of CTNNBL1 generated by MusiteDeep

Figure 3. Predicted post-translational modification sites shown in CTNNBL1 primary structure. P: Phosphorylation; gl: Glycosylation; ub: Ubiquitination; su: SUMOylation; ac: Acetylation; me: Methylation; pc: Pyrrolidone carboxylic acid

Table 4. The characteristics of mutant and wild-type amino acids retrieved from Project HOPE modeling

SNP ID	Amino Acid Change	Mutant Amino Acids			Wild-type Amino Acids		
		Charge	Size	Hydrophobicity	Charge	Size	Hydrophobicity
rs141919968	R323W	Neutral	>	>	Positive	<	<
rs144576870	R73Q	Neutral	<		Positive	>	
rs151227978	R458H	Neutral	<		Positive	>	
rs200219582	P274H		>	<		<	>
rs200874158	K365N	Neutral	<		Positive	>	
rs369990971	N299S		<	>		>	<
rs370193565	R517W	Neutral	>	>	Positive	<	<
rs370578340	R202W	Neutral	>	>	Positive	<	<

Table 5. 3D modeling of human CTNNBL1 protein illustrated by Project HOPE software

A) rs141919968 **R323W** 	**B) rs144576870** **R73Q** 	**C) rs151227978** **R458H** 	**D) rs200219582** **P274H** 
E) rs200874158 **K365N** 	**F) rs369990971** **N299S** 	**G) rs370193565** **R517W** 	**H) rs370578340** **R202W** 

studies. Therefore, it was aimed to conduct an *in silico* study to identify the effects of the high-risk nsSNPs of obesity-associated *CTNNBL1* gene.

Protein stability is vital for structure and correct function of proteins. Protein stability directly affects 3D-structure of proteins and consequently designates their functions. Protein stability changes can lead to protein misfolding, aggregation, and degradation. Moreover, phylogenetic conservation in a protein sequence is crucial to determine the negative effects of a point mutation (23). Post-translational modification that occurs after translation is a chemical modification and contributes to the maturation of protein by structurally and functionally. These PTMs such as phosphorylation and methylation are involved in numerous biological processes such as localization of protein in cell, regulation of protein, and interaction with biomolecules such as DNA, RNA, and lipids. Furthermore, PTMs are significant for protein degradation, folding, and gene and protein expression (24). The disturbed PTMs by nsSNPs can modify the protein structure and function and hence, they may have the potential for being molecular marker candidates (25). Our findings showed that rs141919968, rs144576870, rs151227978, rs200219582, rs200874158, rs369990971, rs370193565, and rs370578340 polymorphisms decreased the stability of beta-catenin-like protein 1. Our study revealed that R323W (rs141919968), R73Q (rs144576870), R458H (rs151227978), P274H (rs200219582), and N299S (rs369990971) were highly conserved and these polymorphisms were damaging to the human beta-catenin-like protein 1. Conversely, no PTMs were determined at residues related with high-risk deleterious nsSNPs.

Based on the web-based bioinformatic tools, we demonstrated five remarkable nsSNPs R323W (rs141919968), R73Q (rs144576870), R458H (rs151227978), P274H (rs200219582), and N299S (rs369990971) from 8 *CTNNBL1* deleterious nsSNPs to focus on since: a) the five missense SNPs were determind to be damaging by at least by four methods; b) these nsSNPs may cause to $\Delta\Delta G < 0$, eventually induce protein stabiliy decrease; c) it was demonstrated that these five nsSNPs were in highly conserved regions; d structural modeling indicated that these four nsSNPs may lead to alterations in hydrophobicity, charge, and size of human beta-catenin-like protein 1.

To elucidate the effects of amino acid substitutions on protein structure and cellular function is significant for determination of the relationship between point mutations and the pathogenesis of human diseases. nsSNPs which result in residue alteration at critical sites in protein may cause several conformational cahnges such as salt bridge cleavage, hydrogen bond disruption, disturbances in interaction network. These alterations may induce protein destabilisation and aggregation and impact protein folding kinetics. It is known that disturbances in protein stability is the common result of disease-associated mutations in various human diseases (26, 27).

Conclusion

SNPs that can be used as molecular markers may be implicated in disease pathogenesis and may predict treatment. In this regard, identification of deleterious nsSNPs has importance to clarify the molecular mechanisms associated with human diseases. The current study investigated the impact of functional nsSNPs in obesity-associated *CTNNBL1* gene by computational methods and only 5 nsSNPs were found to be high-risk deleterious. No study conducted with these deleterious nsSNPs is reported in the literature. This current study enables a guideline for researchers to figure out the roles of these high-risk nsSNPs in the pathogenesis of obesity. This bioinformatics study provides a comprehensive perspective for genetics mechanism of obesity caused by human *CTNNBL1* gene mutations. Prediction deleterious nsSNPs by multiple algorithmic web tools may be advantageous in decreasing time and cost. We suppose that the predicted pathogenic nsSNPs may be implicated in the pathogenesis of obesity and confirmation of the results by population studies are required.

Conflicts of interest : The authors declare that they have no conflict of interest

References

1. Melo LC, Silva MA, Calles AC. Obesity and lung function: a systematic review. Einstein 2014; 12: 120-5. https://doi.org/10.1590/S1679-45082014RW2691

2. Avsar O, Kuskucu A, Sancak S, Genc E. Are dopaminergic genotypes risk factors for eating behavior and obesity in adults? Neurosci Lett 2017; 654: 28-32. https://doi.org/10.1016/j.neulet.2017.06.023
3. Avsar O, Kuskucu A, Sancak S, Genc E. Do Vesicular Monoamine Transporter 2 Genotypes Relate to Obesity and Eating Behavior? Neuropsychiatry (London) 2017; 7: 1020–1025. https://doi.org/10.4172/Neuropsychiatry.1000310
4. Papassotiropoulos A, Stefanova E, Vogler C, Gschwind K, Ackermann S, Spalek K, et al. A genome-wide survey and functional brain imaging study dentify CTNNBL1 as a memory-related gene. Mol Psychiatry 2013; 18: 255–263. https://doi.org/10.1038/mp.2011.148
5. Kuhny M, Forbes LR, Cakan E, Vega-Loza A, Kostiuk V, Dinesh RK, et al. Disease-associated CTNNBL1 mutation impairs somatic hypermutation by decreasing nuclear AID. J Clin Investig 2020; 130: 4411-4422. https://doi.org/10.1172/JCI131297
6. Liu YJ, Liu XG, Wang L, Dina C, Yan H, Liu JF, et al. Genome-wide association scans identified CTNNBL1 as a novel gene for obesity. Hum Mol Genet 2008; 17: 1803–1813. https://doi.org/10.1093/hmg/ddn072
7. Moon RT, Bowerman B, Boutros M, Perrimon N. The promise and perils of Wnt signaling through beta-catenin. Science 2002; 296: 1644–1646. https://doi.org/10.1126/science.1071549
8. Ross SE, Hemati N, Longo KA, Bennett CN, Lucas PC, Erickson RL, et al. Inhibition of adipogenesis by Wnt signaling. Science 2000; 289: 950–953. https://doi.org/10.1126/science.289.5481.950
9. Vogel C, Greene B, Scherag A, Müller TD, Friedel S, Grallert H, et al. Non-replication of an association of CTNNBL1 polymorphisms and obesity in a population of Central European ancestry. BMC Med Genet 2009; 10: 14. https://doi.org/10.1186/1471-2350-10-14
10. Andreasen CH, Mogensen MS, Johnsen KB, Sandbæk A, Lauritzen T, Almind K, et al. Studies of CTNNBL1 and FDFT1 variants and measures of obesity: analyses of quantitative traits and case-control studies in 18,014 Danes. BMC Med Gen 2009; 10: 17. https://doi.org/10.1186/1471-2350-10-17
11. Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. Bioinformatics 2016; 32: 2230–2232. https://doi.org/10.1093/bioinformatics/btw222
12. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015; 31: 2745–2747. https://doi.org/10.1093/bioinformatics/btv195
13. Capriotti E, Calabrese E, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006; 22: 2729–2734. https://doi.org/10.1093/bioinformatics/btl423
14. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Pauline C. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 2012; 40: W452–W457. https://doi.org/10.1093/nar/gks539
15. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. BMC Genomics 2015; 16: S1. https://doi.org/10.1186/1471-2164-16-S8-S1
16. Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. Bioinformatics 2005; 21: ii54-8. https://doi.org/10.1093/bioinformatics/bti1109
17. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 2006; 62: 1125–1132. https://doi.org/10.1002/prot.20810
18. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 2016; 44: W344–W350. https://doi.org/10.1093/nar/gkw408
19. Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S, et al. MusiteDeep: a deep learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Res 2020; 48: W140–W146. https://doi.org/10.1093/nar/gkaa275
20. Venselaar H, Beek T, Kuipers R, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics 2010; 11: 548. https://doi.org/10.1186/1471-2105-11-548
21. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, et al. ConSurf: using evolutionary data to raise testable hypotheses about protein function. Isr J Chem 2013; 53: 199–206. https://doi.org/10.1002/ijch.201200096
22. Morze J, Rücker G, Danielewicz A, Przybyłowicz K, Neuenschwander M, Schlesinger S, et al. Impact of different training modalities on anthropometric outcomes in patients with obesity: A systematic review and network meta-analysis. Obes Rev 2021; e13218. https://doi.org/10.1111/obr.13218
23. Hossain S, Roy AS, Islam S. In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. Sci Rep 2020; 10: 14542. https://doi.org/10.1038/s41598-020-71457-1
24. Tak I, Ali F, Dar JS, Magray AR, Ganai BA, Chisti MZ. Posttranslational modifications of proteins and their role in biological processes and associated diseases. Protein Modificomics 2019; pp. 1-35.
25. Subbiah HV, Babu PR, Subbiah U. In silico analysis of non-synonymous single nucleotide polymorphisms of human DEFB1 gene. Egypt J Med Hum Genet 2020; 21: 66. https://doi.org/10.1186/s43042-020-00110-3
26. Pauline C, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics

Hum Genet 2006; 7: 61-80. https://doi.org/10.1146/annurev.genom.7.080505.115630
27. Teng S, Wang L, Srivastava AK, Schwartz CE, Alexov E. Structural assessment of the effects of amino acid substitutions on protein stability and protein-protein interaction. IJCBDD 2010; 3*:* 334–349. https://doi.org/10.1504/IJCBDD.2010.038396

Correspondence
Orcun Avsar,
Hitit University,
Department of Molecular Biology and Genetics,
Corum, Turkey
E-mail: orcunavsar@hitit.edu.tr