BioBacta

Journal of Bioscience and Applied Research
www.jbaar.org

In Silico approach for identification, prediction of AMPD1 gene nsSNPs associated with Myoadenylate Deaminase deficiency

Abrar Mohamed Gamar Mohamed[1,2*], Abdelrahman Hamza Abdelmoneim Hamza[2,3], Hiba Awadelkareem Osman Fadl[4,5], Afra Mohamed Suliman Albkrye[6], Hadeel Abdelsamea Mohamed Ahmed[7], Hazem Abdo Mohamed Abubaker[8] and Sahar Gamal Elbager[9,10]

[1*]Faculty of Medicine, Al-Zaiem Al-Azhari University, Khartoum, Sudan. abrar.gamer94@gmail.com
[2]Clinical Immunology Resident, Sudan Medical Specialization Board, Khartoum, Sudan.
[3]Faculty of Medicine, Al-Neelain University, Khartoum, Sudan. abduhamza009@gmail.com
[4]Department of Haematology, Faculty of Medical Laboratory Sciences, Al-Neelain University, Khartoum, Sudan. heba2015@hotmail.com
[5]Department of Medical Laboratory, Sudanese Medical Research Association, Khartoum, Sudan.
[6]Department of Molecular Biology and Bioinformatics, Faculty of Veterinary Medicine, Bahri University, Khartoum, Sudan. aframoh2016@bahri.edu.sd
[7]Department of Biotechnology, Faculty of Science and Technology, Omdurman Islamic University, Khartoum, Sudan. hadeelabdelsamea@gmail.com
[8]Department of Clinical Medicine, Faculty of Veterinary Medicine, University of Khartoum, Khartoum, Sudan. mr.haziem@hotmail.com
[9]Department of Haematology, Faculty of Medical Laboratory Sciences, University of Medical Sciences and Technology, Khartoum, Sudan. saharelbagir@gmail.com
[10]Department of Pathology, South Egypt Cancer Institute, Assiut University, Assiut, Egypt.

DOI: 10.21608/jbaar.2023.284727

Abstract

Background: Myoadenylate deaminase deficiency is an autosomal recessive metabolic myopathy caused by mutations in the Adenosine monophosphate deaminase 1 gene. Adenosine monophosphate deaminase 1 gene deficiency is one of the most common causes of exercise-induced myopathy. In this study, non-synonymous single nucleotide polymorphism was analyzed for its functional and structural impact which is deleterious to Adenosine monophosphate deaminase 1 protein. **Methods:** The data on human Adenosine monophosphate deaminase 1gene was retrieved from the NCBI database on 9 JUNE 2021 and then analyzed using different bioinformatics prediction algorithms, namely: SIFT, PolyPhen-2, PROVEAN, SNAP2, PANTHER, SNPs and GO, PMut, and I-Mutant to detect the deleterious nsSNPs and its association with diseases. In addition, a Consurf web server was used to detect the functional SNPs in the conserved region. Chimera, Project Hope, and MutPred2 software were used to visualize and analyze the effect of nsSNPs on the functions and structure of the AMPD1 protein. Finally, both the STRING database and KEGG were used for the prediction of protein-protein interaction. **Results:** A total of 6178 SNPs were reported in the human AMPD1 gene. In this study 583 missense nsSNPs were selected for investigation and only 72 nsSNPs were shortlisted and computationally evaluated for their impact on AMPD1 protein. From all servers that were used collectively (K320I, R421W, R458C, R458H, R51C, R757L, R761H, and G246S) nsSNPs were predicted as deleterious, associated with disease, highly conserved, and decrease effective stability of AMPD1 protein. In addition, the AMPD1 protein was predicted to have strong interactions with ten proteins involved in various ranges of biological processes.

Conclusion: The present study undertakes a systematic bioinformatics approach to identify functionally important nsSNPs in the human AMPD1 gene to understand how these mutations affect the protein function and structure and hence promote a myoadenylate deaminase deficiency.

Keywords: *AMPD1 Gene, Myoadenylate Deaminase Deficiency, Myopathy, Bioinformatics, nsSNPs, & In Silico.*

Received: October 22, 2022. Accepted: December 25, 2022. Article in Press: February 9, 2023. Published: March 1, 2023

1 Introduction

Myoadenylate deaminase deficiency (MAD) is an autosomal recessive metabolic myopathy caused by mutations in the Adenosine monophosphate deaminase 1(AMPD1) gene. It is most commonly associated with exercise-induced myalgia, rhabdomyolysis, infantile hypotonia, or an increase in the level of creatine kinase. Although, people with MAD may not exhibit any symptoms. There is a common polymorphism of the nonsense mutation with a C-to-T transition at position 34 of the AMPD1 gene in most cases of Myoadenylate deaminase deficiency, which results in a premature stop codon resulting in some metabolic muscle diseases **(Feng et al., 2017)**.

AMPD1 gene is one of three members of the AMPD multigene family, which is localized in chromosome 1 in the region p13.2 with approximately 20 kilobases in size including 16 exons, and 15 introns **(Sabina et al., 1990)**.

The AMPD1 gene encodes a crucial enzyme called adenosine monophosphate deaminase (AMPD). AMPD enzyme is predominantly expressed in skeletal muscles, where it plays a role in producing energy. Specifically, during physical activity, as a part of a process called the purine nucleotide cycle. This cycle uses purine molecules, which are a group of building blocks of DNA and RNA nucleotides, and molecules such as AMP that serve as energy sources in the cell. As part of the purine nucleotide cycle, AMP deaminase converts AMP into IMP with the release of ammonia and as the cycle continues, molecules are produced can be used by muscle cells for energy **(Ahmetov and Fedotovskaya, 2015)**.

A deficiency of the AMPD1 gene is one of the most common causes of exercise-induced myopathy and may be the cause of metabolic myopathy among human beings. In Caucasian individuals, AMPD1 deficiency is considered one of the most common inherited defects, with several studies indicating that 2–3% of Caucasians are homozygous for AMPD1 defects **(Cheng et al., 2014)**.

The first report of primary myoadenylate deaminase deficiency was from a Japanese patient with progressive weakness and muscle atrophy, her DNA analysis revealed compound heterozygous missense mutations of the AMPD1 gene **(Abe et al., 2000)**. Another case study done for a patient with severe skeletal muscle pain and increasing creatine kinase, MAD, was confirmed by histochemistry and molecular testing, which revealed that recurrent homozygous pathogenic variants of the AMPD1 gene are present in most patients with MAD **(Lim et al., 2017)**. In a mouse model study, it has been found that decreased AMPD1 expression reduced muscle strength in healthy mice, and that AMPD1 deficiency was responsible for the observed muscle weakness in mouse model myositis **(Coley et al., 2012)**.

The human genome project revealed that 99.5% of the human genome is identical worldwide and the rest 0.5% of the genome contains individual variations **(Hood and Rowen, 2013)**. One of the most frequent types of human DNA variants is single nucleotide polymorphisms, therefore at least 10 million SNPs were found within the genome occurring approximately every 100-300 base pairs **(Robert and Pelletier, 2018)**. The most significant SNPs were found in the coding regions, which were divided into synonymous and non-synonymous (nsSNPs). While synonymous mutation is thought to be functionally silent and evolutionary neutral, non-synonymous mutation may change the sequence of a protein **(Chu and Wei, 2019)**. However, some nsSNPs may cause a damaging effect on the protein while others considered to be neutral. Thus, an effective computational approach is imperative to predict the impact of nsSNPs in the protein.

The AMPD1 gene and myoadenylate deaminase deficiency have been detected through different types of studies, but there hasn't been a comprehensive bioinformatics analysis of the human AMPD1 gene and its effect on the structure and function of AMPD1 protein, which may have an important role in disease susceptibility. Bioinformatics is a science field that uses DNA and amino acid sequences to solve biological problems and related information. Therefore, Various bioinformatics tools are incorporated into biological research for storing, retrieving, analyzing, annotating, and visualizing results **(Wani. et al., 2018)**. Hence, it is low-cost, affordable, and involves a short time frame. In this study, by using different bioinformatics tools we aimed to understand the effect of multiple nsSNPs of AMPD1 gene and it is deleterious effect on AMPD1 protein, which further will enable better diagnosis, treatment, and preventative measures to be developed.

2 Material and Methods

2.1 Data Set

The data on the human AMPD1 gene (Gene ID: 270) was retrieved from the Entrez Gene database from National Center for Biological Information (NCBI) database on 9 JUNE 2021. The AMPD1 protein sequence (accession ID: P23109) and SNPs information of the AMPD1 gene were obtained from UniProtKB databases (http://www.uniprot.org) and NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp/) respectively.

2.2 Identification of Deleterious nsSNPs

The deleterious missense nsSNPs associated with the AMPD1 gene were chosen because they may disrupt the amino acid sequence, thus potentially impairing the function and the structure of the protein. To determine the effect of the nsSNPs mutations on the functions of the protein and whether it will have a deleterious, damaging, or neutral effect, the AMPD1 ns SNPs were analyzed using four different tools (SIFT, PolyPhen -2, PROVEAN, and SNAP2). In addition to nsSNPs predicted to be deleterious by these four tools, were categorized as high-risk nsSNPs and subjected to further analysis like associations with diseases, stability analysis, and structural analysis.

2.2.1 SIFT (Sorting **I**ntolerant **from T**olerant; http://siftdna. org/www/ SIFT_dbSNP.html) server was used to predict whether an amino acid substitution affects protein function based on the degree of conservation of amino acids and so changes at well-conserved positions tend to be predicted as deleterious (15). Using sequence homology substitution and amino acid properties, SIFT predicts a probability below 0.05 to be deleterious and intolerant, whereas a probability greater or equal to 0.05 is considered tolerant **(Ng and Henikoff, 2001)**. nsSNPs retrieved from dbSNP were used as input to SIFT.

2.2.2 PolyPhen -2 (**poly**morphism **phen**otyping v2; http://genetics.bwh.harvard.edu/pph2/) PolyPhen -2 is an automatic tool that was used to predict the possible impact of nsSNPs on the structure and function of a human protein using sequence-based characterization. It extracts protein sequence annotations and structural properties and creates conservation profiles based on single-nucleotide polymorphisms (SNPs) **(Adzhubei et al., 2010)**. According to PolyPhen-2 output prediction, the substitution is either probably harmful, possibly damaging, or benign. UniProtKB database with accession ID P23109 along with the position and name of the wild type was used as a reference source in PolyPhen-2 for all protein sequences and annotations.

2.2.3 PROVEAN (**Pro**tein **V**ariation **E**ffect **An**alyzer ;http://provean.jcvi.org) the nsSNPs identified to be damaged by both SIFT and PolyPhen-2 were then subjected to PROVEAN. In this software, amino acid substitutions and indels are predicted to have some effect on a protein function. The final score under the threshold of −2.5 was considered deleterious, while scores above it were considered neutral **(Choi and Chan, 2015)**. The used input query was a protein FASTA sequence along with amino acid substitutions and it analyses the nsSNPs as deleterious or neutral.

2.2.4 SNAP2 (**S**creening of **N**on-**A**cceptable **P**olymorphism 2; https://rostlab.org/services/snap/) is a tool for Predicting the functional effects of sequence variants based on neural network. The input query submitted was the protein FASTA sequence, it classifies nsSNPs into neutral or damaging according to provided scores of each substitution; the higher the score the more reliable the prediction **(Hecht et al., 2013)**.

2.3 Prediction of disease-associated SNPs

2.3.1 PANTHER (Protein **AN**alysis **TH**rough **E**volutionary **R**elationships**,** http://www.pantherdb.org) is a software that predicts whether a given nsSNP will have a functional impact on a protein, with the cutoff score − 3 indicating a deleterious substitution **(Mi et al., 2016)**. The used input query was a reference protein sequence along with amino acid substitutions and it analyses the nsSNPs as either deleterious or neutral.

2.3.2 SNPs and GO (http://snps-and-go.biocomp.unibo.it/snps-and-go/) a server was used to predict single point mutations of the protein likely to cause disease in humans, using information derived from its sequence, phylogenetic relation, and function. Currently, SNPs and GO are considered one of the best algorithms for predicting deleterious SNPs **(Capriotti et al., 2013)**. The input used was protein FASTA sequence along with amino acid substitution and it analyses the nsSNPs as disease-associated or neutral.

2.3.3 PMut (http://mmb2.pcb.ub.es:8080/PMut) is a web server that was used for the illustration of disease-associated protein mutations. It allows the fast and accurate prediction of the pathological character of single-point amino acid mutations. UniProt accession ID of AMPD1 (P23109) along with the name and position of the wild type and the mutant amino acid was submitted as input for this server. The final output is displayed as a pathogenicity index ranging from 0 to 1 (indexes > 0.5 single pathological mutations) and a confidence index ranging from 0 (low) to 9 (high) **(Ferrer-Costa et al., 2005)**.

2.4 Prediction of nsSNPs impact on the protein stability by I-Mutant2.0 (http://gpcr.biocomp.unibo.

it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi) is a tool that predicts the change in stability of a protein caused by single point mutations in this server the input method either to be in the form of protein sequence or structure **(Capriotti et al., 2005)**. For the present study, input was submitted in the form of a protein FASTA sequence.

2.5 Identification of functional SNPs in the conserved regions by conSurf server (https://consurf.tau.ac.il/) is a tool for revealing functional regions in macromolecules by analyzing the evolution of amino acid and nucleic acid substitutions on homologous sequences. It can predict protein-active sites and identify biologically active peptides **(Ashkenazy et al., 2010)**. Consurf gives an output in the form of a score range from 1 to 9, where score 9 represents the most conserved amino acid and score 1 represents the most variable amino acid. The protein sequence in FASTA format served as the input to ConSurf.

2.6 Analyzing the Effect of nsSNPs on 3D Structure of the Proteins and Physiochemical Properties

2.6.1 Chimera server (www.rbvi.ucsf.edu/chimera/) Is a program that provides interactive visualization and analysis of molecular structures and related data, such as density maps, trajectories, and sequence alignment. We introduce the reference protein sequence to the RaptorX server to generate a structure file of our protein. Then we visualized the AMPD1 structure using chimera software version 1.10.2 **(Pettersen et al., 2004)**.

2.6.2 Project HOPE *(*https://www3.cmbi.umcn.nl/hope/input*/)* is a tool that Provides an analysis of the structural effects of a point mutation in a protein sequence. FASTA sequence of whole protein and selection of mutant variants were submitted to the Project HOPE server to be an input method and the output depended on the structural variation between mutant and wild-type residues **(Venselaar et al., 2010)**.

2.6.3 Mutpred2 (http://mutpred.mutdb.org/) is a tool used for predicting the effects of coding variants on structural, functional, and phenotypic traits **(Li et al., 2009)**. In particular, it attempts to infer molecular mechanisms of disease from protein-coding variants by predicting probabilities of gain or loss of structure and function. It has a numerical scoring system called general score (g) with a 0–1 value range; variants with a score > 0.5 are considered as potentially harmful to protein function and those with a score >0.75 as high confidence harmful changes. The input was submitted in the form of protein sequences along with amino acid substitutions.

2.7 Predictions of protein-protein interactions

2.7.1 STRING (**S**earch **T**ool for the **R**etrieval of **I**nteracting **G**enes/proteins; https://string-db.org) is a database that provides an assessment and integration of direct and indirect interactions between proteins **(von Mering et al., 2005)**. The input used is the protein name and the organism.

2.7.2 KEGG (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/) is a database resource that can be used to analyze high-level functions of the biological system. It can analyze gene functions to understand metabolic pathways, regulatory pathways, and molecular complexes associated with biological systems **(Kanehisa et al., 2017)**.

3 Results

3.1 SNPs Dataset

The nsSNPs of the AMPD1 gene examined in this study were retrieved from the NCBI dbSNP database. A total of 6178 SNPs were reported in the human AMPD1 gene. Among reported SNPs, 612 non-synonymous SNPs (583 missenses, 29 nonsense), 216 coding synonymous SNPs, 4445 SNPs in the intron region,16 SNPs in 5' UTR region, 2 SNPs in the 3' UTR region, and the rest were other types.

In this study, the 583 missense nsSNPs were selected for investigation since deleterious nsSNPs could have structural and functional impacts on the protein. The missense nsSNPs were sorted out on basis of their minor allele frequency MAF range (0.001–0.05), and only 72 SNPs were shortlisted and computationally evaluated for their impact on AMPD1 protein using different bioinformatics tools.

3.2 Prediction of deleterious nsSNPs of AMPD1 gene

A total number of 72 nsSNPs on the AMPD1 gene were submitted as a batch to the SIFT program. According to our SIFT analysis, 20 nsSNPs were predicted to be deleterious, which were submitted for further analysis by Polyphen-2, PROVEAN, and SNAP2, their results were 15 nsSNPs probably damaging, 16 nsSNPs were deleterious and 18 nsSNPs have affected the function of the protein, respectively as shown in (Table.1). Therefore, out of 20 nsSNPs, 15 nsSNPs were classified as deleterious by Polyphen-2, PROVEAN, and SNAP2, then taken into consideration and further analyzed by other tools.

Table 1: List of nsSNPs analysis by SIFT, PolyPhen-2, PROVEAN, and SNAP2

SNP	Amino acid change	SIFT prediction	PolyPhen-2 prediction	PROVEAN prediction (cutoff=-2.5)	SNAP-2 Prediction
rs2273268	E55K	Deleterious	Benign	Neutral	Effect
rs374697989	E714Q	Deleterious	Probably damaging	Neutral	Effect
rs373262677	**G246S**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs150645738	**G607E**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs182610890	**I677N**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs34526199	**K320I**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs61752478	**M343I**	**Deleterious**	**Possibly damaging**	**Deleterious**	**Effect**
rs150886824	**M489T**	**Deleterious**	**Possibly damaging**	**Deleterious**	**Effect**
rs61752479	**P81L**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs12566550	R108C	Deleterious	Probably damaging	Neutral	Neutral
rs142123340	R350H	Deleterious	BENIGN	Deleterious	Effect
rs143583425	**R399C**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs35859650	**R421W**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs187527797	**R458C**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs121912682	**R458H**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs369335745	**R51C**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs143596876	**R749W**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs146036570	**R757L**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs142643298	**R761H**	**Deleterious**	**Probably damaging**	**Deleterious**	**Effect**
rs145315039	S100F	Deleterious	Benign	Neutral	Neutral

3.3 Prediction of Disease Associated SNPs

PANTHER, SNPs and GO, and PMut were used to predict the results obtained from four previous tools. Out of 15 nsSNPs that were predicted to be deleterious previously; PANTHER predicted 13 nsSNPs were associated with disease, while SNPs and GO and PMut predicted all 15 nsSNP to be associated with disease (Table. 2). Finally, out of 15 nsSNP, 13 nsSNPs were found to be associated with disease by all three tools were used. Which was then taken into consideration and further analyzed by other tools.

3.4 Prediction of nsSNPs impact on the Protein Stability

All SNPs predicted to be associated with the disease were submitted to I-mutant 2.0 server. The results were as follows: out of 13 nsSNPs predicted from previous programs entered, I-mutant 2 server predicted 10 nsSNPs were decreased effective stability of the AMPD1 protein, result listed in (Table. 3).

3.5 Identification of functional SNPs in the conserved regions

The input to the ConSurf server was 13 nsSNPs disease-associated and decrease effective stability of the AMPD1 protein predicted by previous tools. ConSurf predicted that 11 nsSNPs were highly conserved; from which (G607E, K320I, R421W, R458C, R458H, and R761H) are exposed and have a functional effect, (G246S, I677N, and R757L) are buried and have structural effects, (P81L) are exposed, and (R51C) has no structural or functional effects on the AMPD1 protein.

The other two nsSNPs (R749W and R399C) have low conservation of the protein. The results are shown in (Fig. 1).

Finally, out of 13 nsSNPs analyzed by both I-mutant and ConSurf software eight nsSNPs namely (G246S, K320I, R421W, R458C, R458H, R51C, R757L, and R761H) were predicted to be highly conserved and decrease effective stability of the AMPD1 protein, which was then taken into further analysis.

3.6 Analyzing the Effect of nsSNPs on the 3D structure of the protein and physiochemical properties using Chimera, Project HOPE, and MutPred2 server

From eight AMPD1 nsSNPs predicted after comparing results of all previously mentioned tools as disease-associated; Chimera software visualized that all of these nsSNPS entered were mutant (Figure. 3). While project HOPE server predicted the common physiochemical consequence of AMPD1 nsSNPs which resulted in variations in amino acid properties such as charge, size, flexibility, and hydrophobicity value between wild-type and mutant residues as shown in (Table. 4). However, one of the changes in physiochemical properties showed that (R421W, R458C, R458H, and R757L) nsSNPs their mutant residues are highly conserved and their substitutions are probably damaging to the protein structure. Although, all AMPD1 nsSNPs are predicted to be present mainly in Amp Deaminase 1 'IPR029770' domain and they have shared molecular functions such as catalytic, hydrolase, and deaminase activities.

MutPred2 results showed that (R51C) nsSNPs were harmful, While the rest of nsSNPS were highly harmful. The top possible molecular mechanisms disrupted were predicted such as Altered ordered or disordered interface, Loss of Intrinsic disorder, loss of Acetylation, Loss of the allosteric site, and gain of the strand. Table. 5 summarizes the result obtained from the MutPred2 server.

3.7 Protein-protein interaction analysis

STRING software identified that AMPD1 protein has strong interactions with ten proteins including, ADK, ADSS, ADSSL1, ADSL, APRT, ATIC, ITPA, HPRT1, GMPR, and IMPDH2 as shown in (Figure 3). Furthermore, the STRING analysis identified that this set of proteins is involved in various range of biological processes such as metabolic/biosynthetic processes of nucleobase-containing small molecules, nucleoside phosphate, purine nucleotide, and carbohydrate derivative (figure 4). In molecular interaction analysis, AMPD1 was predicted to be directly involved in catalytic activity, ion binding, nucleotide binding, hydrolase activity, and ATP binding (figure. 5). Additionally, some pathways important in nucleotide metabolism, metabolic pathways, and drug metabolism were also enriched (figure 6). KEGG analysis reveals that AMPD1 is involved in the purine nucleotide pathway and many metabolic pathways as shown in (Figure 7).

Table. 2: List of nsSNPs predicted as a disease associated by PANTHER, SNPs and GO and PMut

SNP	Amino acid change	PANTHAR	SNP&go	PMut
rs373262677	G246S	**Disease**	**Disease**	**Disease**
rs150645738	G607E	**Disease**	**Disease**	**Disease**
rs182610890	I677N	**Disease**	**Disease**	**Disease**
rs34526199	K320I	**Disease**	**Disease**	**Disease**
rs61752478	M343I	Neutral	Disease	Disease
rs150886824	M489T	Neutral	Disease	Disease
rs61752479	P81L	**Disease**	**Disease**	**Disease**
rs143583425	R399C	**Disease**	**Disease**	**Disease**
rs35859650	R421W	**Disease**	**Disease**	**Disease**
rs187527797	R458C	**Disease**	**Disease**	**Disease**
rs121912682	R458H	**Disease**	**Disease**	**Disease**
rs369335745	R51C	**Disease**	**Disease**	**Disease**
rs143596876	R749W	**Disease**	**Disease**	**Disease**
rs146036570	R757L	**Disease**	**Disease**	**Disease**
rs142643298	R761H	**Disease**	**Disease**	**Disease**

Table 3: Prediction of protein stability by I-Mutant

SNP ID	Amino acid change	Prediction	RI	PH	Temp
rs373262677	G246S	Decrease	9	7	25
rs150645738	G607E	Increase	1	7	25
rs182610890	I677N	Increase	1	7	25
rs34526199	K320I	Decrease	5	7	25
rs61752479	P81L	Increase	2	7	25
rs143583425	R399C	Decrease	5	7	25
rs35859650	R421W	Decrease	6	7	25
rs187527797	R458C	Decrease	4	7	25
rs121912682	R458H	Decrease	9	7	25
rs369335745	R51C	Decrease	3	7	25
rs143596876	R749W	Decrease	5	7	25
rs146036570	R757L	Decrease	4	7	25
rs142643298	R761H	Decrease	8	7	25

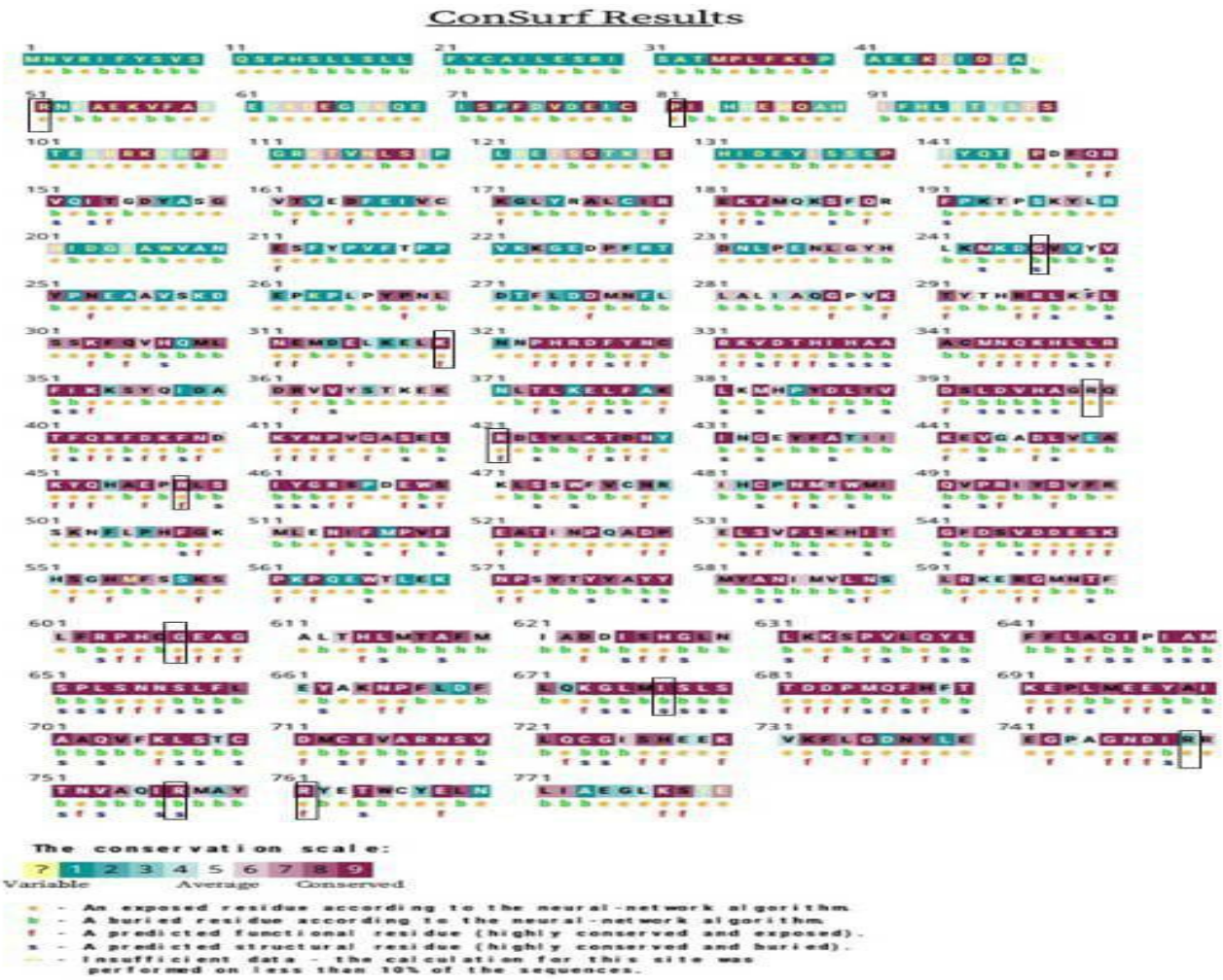


Figure 1. Analysis of evolutionarily conserved amino acid residues of AMPD1 protein by ConSurf. The color coding bar shows the coloring scheme representation of the conservation score.

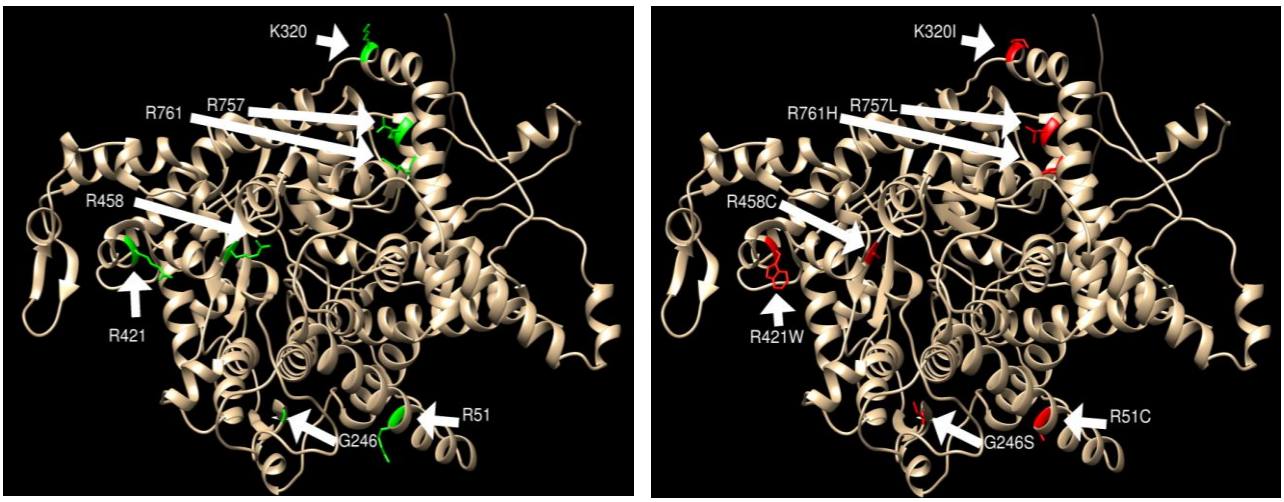


Figure 2: Predicted 3D structure of AMPD1 protein visualized using Chimera software. The green color represents the wild amino acids while the red color represents the mutant amino acid

Table. 4: Schematic structures of the wild-type residue (left) and mutant residue (right) amino acid for each Mutation detected by the project HOPE server

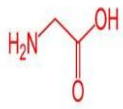
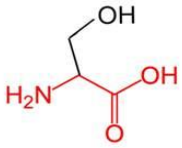
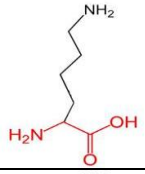
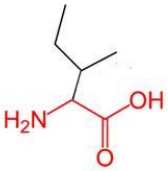
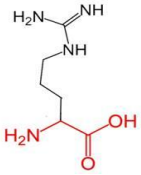
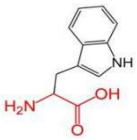
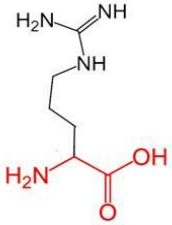
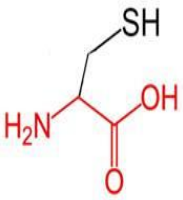
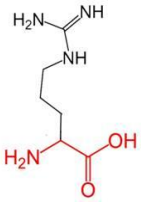
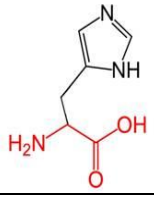
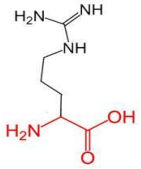
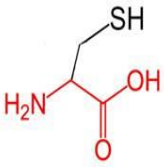
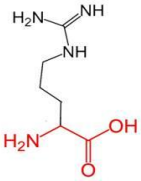
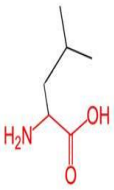
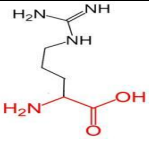
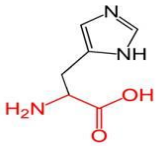
Residue	Structure of wild residue	Structure of mutant residue	Properties
G246S			• The mutant residue is bigger than the wild-type residue. • The wild-type residue is very conserved • Mutant at this position and this mutation is possibly not damaging to the protein.
K320I			• The mutant residue is smaller than the wild-type residue. • The wild-type residue charge was positive; the mutant residue charge is neutral. • The mutant residue is more hydrophobic than the wild-type residue.
R421W			• The mutant residue is bigger than the wild-type residue. • The wild-type residue charge was positive; the mutant residue charge is neutral. • The mutant residue is more hydrophobic than the wild-type residue. • The mutant residue is located near a highly conserved position.
R458C			• The mutant residue is smaller than the wild-type residue. • The wild-type residue charge was positive; the mutant residue charge is neutral. • The mutant residue is more hydrophobic than the wild-type residue. • The mutant residue is located near a highly conserved position.
R458H			• The mutant residue is smaller than the wild-type residue. • The wild-type residue charge was positive; the mutant residue charge is neutral. • The mutant residue is located near a highly conserved position.
R51C			• The mutant residue is smaller than the wild-type residue. • The wild-type residue charge was positive; the mutant residue charge is neutral. • The mutant residue is more hydrophobic than the wild-type residue.
R757L			• The mutant residue is smaller than the wild-type residue. • The wild-type residue charge was positive; the mutant residue charge is neutral. • The mutant residue is more hydrophobic than the wild-type residue. • The mutant residue is located near a highly conserved position.
R761H			• The mutant residue is smaller than the wild-type residue. • The wild-type residue charge was positive; the mutant residue charge is neutral. • This mutation is probably damaging to the protein.

Table. 5: Analysis of the effect of nsSNPs in AMPD1 structure, function, and evolution by MutPred2 server.

SNP ID	Amino acid change	MutPred2 Score(g)	Prediction	Molecular mechanisms with P-values <= 0.05
rs373262677	G246S	0.830	Highly harmful	• Gain of Strand (P =0.03) • Gain of Sulfation at Y249 (P= 0.03)
rs34526199	K320I	0.777	Highly harmful	• Loss of Intrinsic disorder (P =0.05) • Altered Ordered interface (P= 0.02) • Loss of B-factor (P= 0.04) • Altered Metal-binding (P =0.01) • Altered DNA binding (P =0.02) • Loss of Allosteric site at R325(P= 0.04)
rs35859650	R421W	0.916	Highly harmful	• Altered Ordered interface (P =3.0e-03) • Loss of Helix (P= 0.03) • Gain of Relative solvent accessibility (P= 0.03) • Gain of Allosteric site at R421 (P= 0.01) • Altered Transmembrane protein (P= 5.2e-03)
rs187527797	R458C	0.895	Highly harmful	• Altered Disordered interface (P =0.05) • Altered Ordered interface (P= 0.01) • Altered Transmembrane protein (P =0.02)
rs121912682	R458H	0.830	Highly harmful	• Altered Ordered interface (P= 0.01) • Altered Transmembrane protein (P =0.02)
rs369335745	R51C	0.598	Harmful	• Loss of Acetylation at K56 (P =4.8e-03) • Loss of Methylation at K56(P =0.03)
rs146036570	R757L	0.951	Highly harmful	• Loss of Allosteric site at R757 (P =5.4e-05) • Altered Ordered interface (P =1.2e-03) • Altered DNA binding (P =2.0e-03)
rs142643298	R761H	0.892	Highly harmful	• Altered Ordered interface (P=0.01) • Loss of Allosteric site at Y760 (P=7.2e-03) • Altered DNA binding (P=0.02) • Altered Metal-binding (P=0.04)

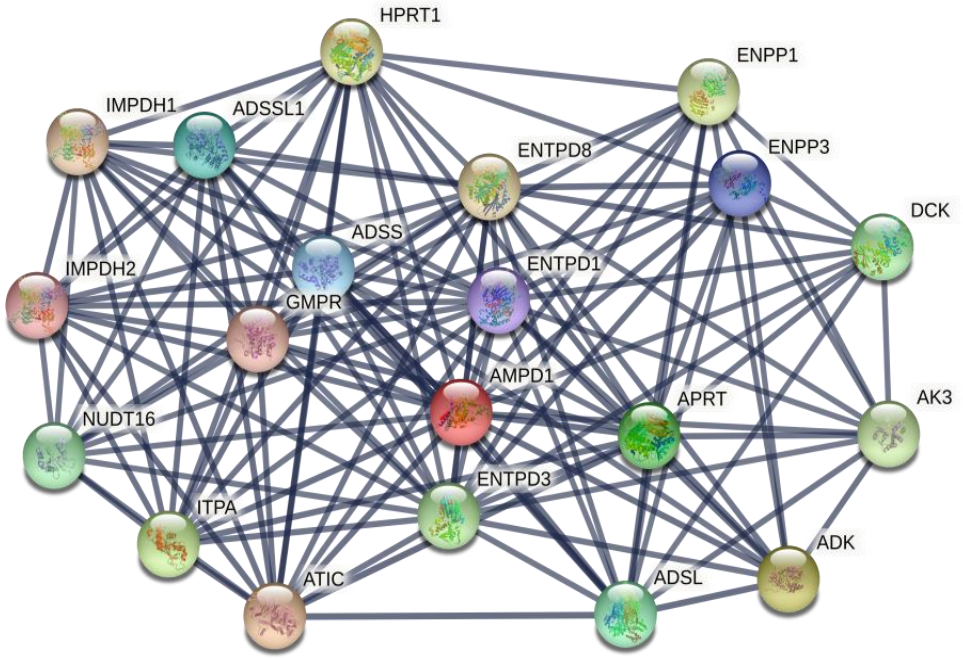


Figure 3. Functional interaction between AMPD1 and it is related genes

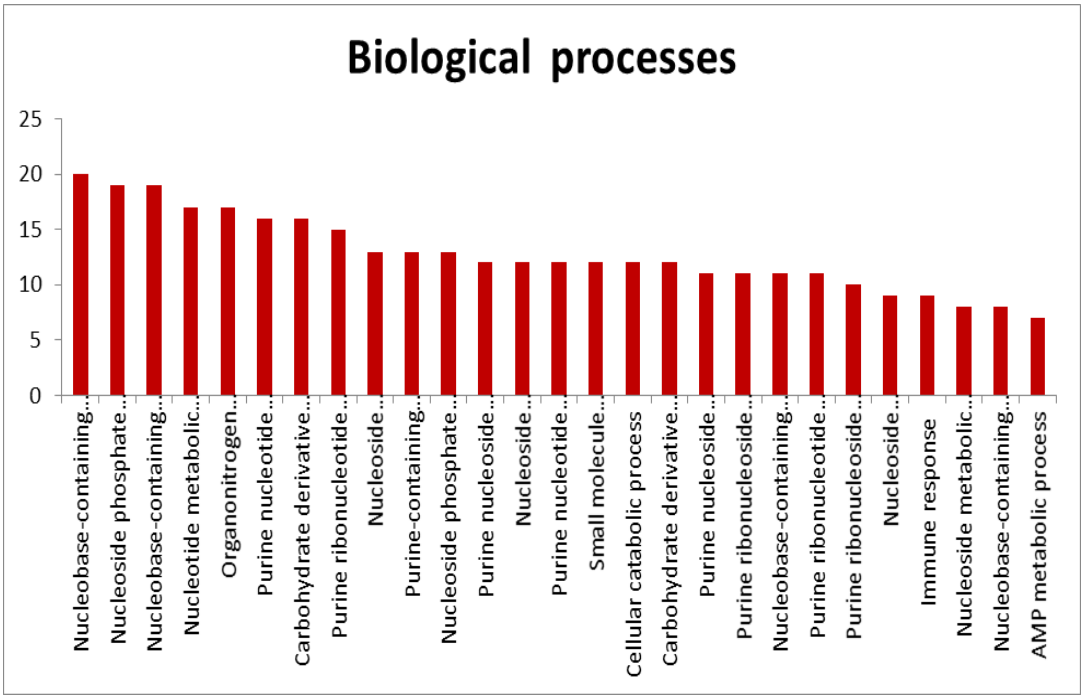


Figure 4. Biological processes of AMPD1 protein detected by STRING software

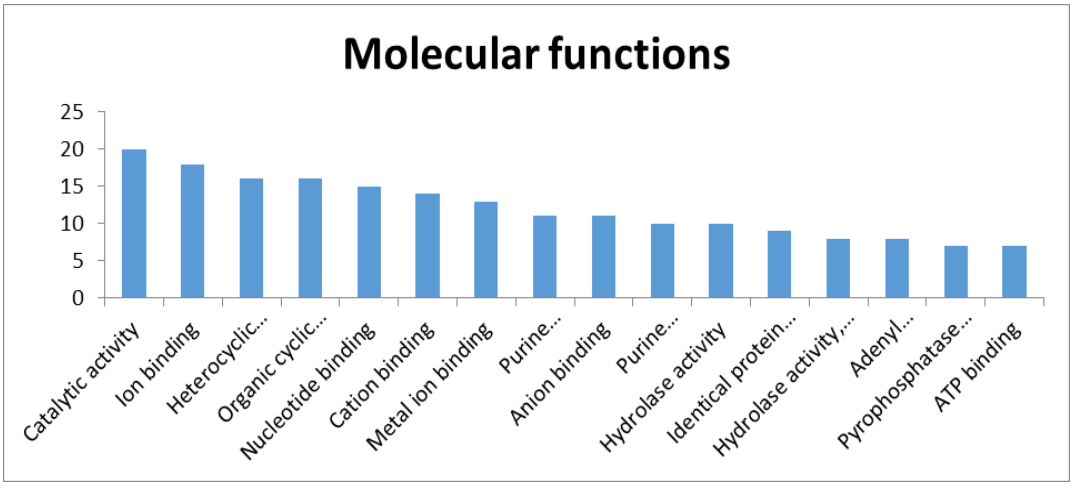


Figure 5. Molecular function analysis of AMPD1 gene using STRING software

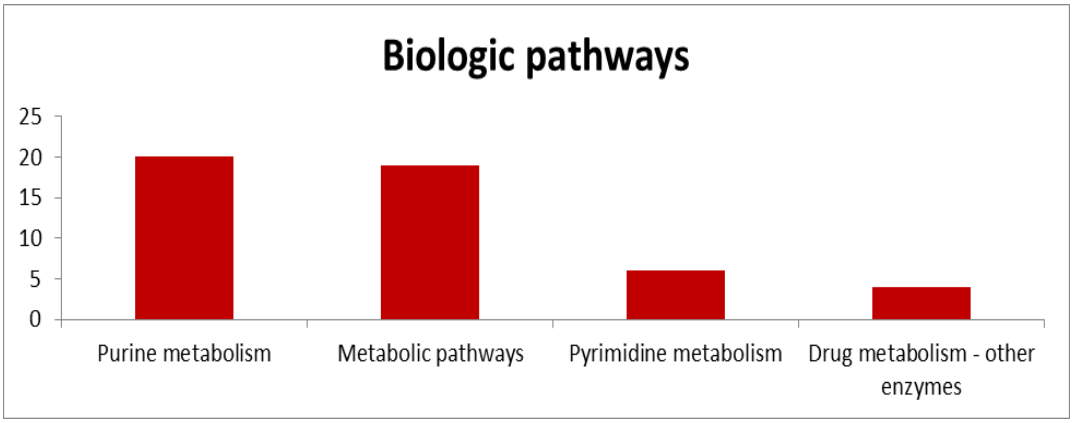


Figure 6. Biological pathways important of AMPD1 protein

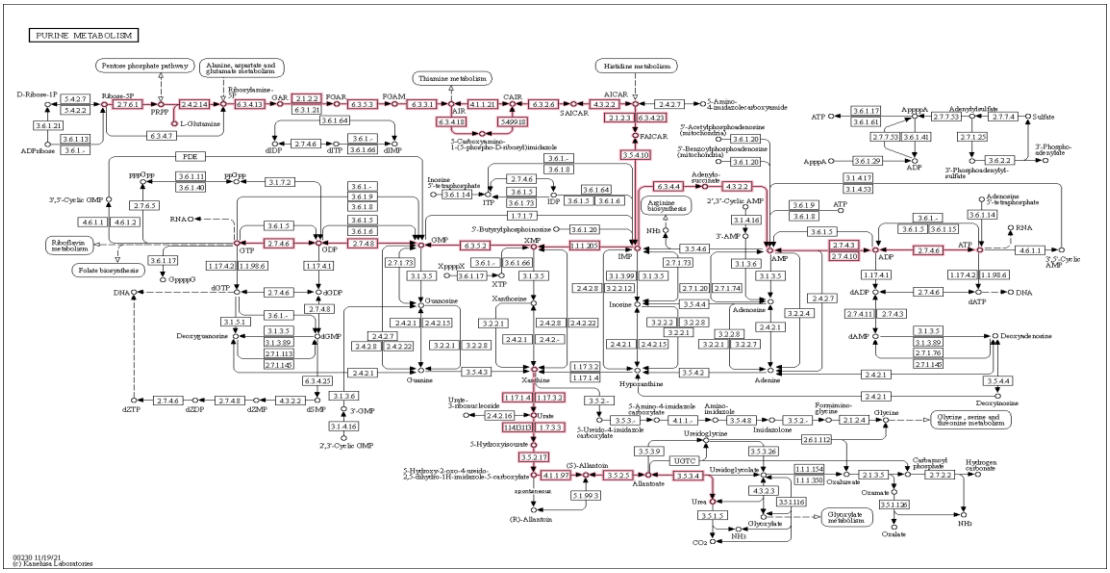


Figure 7. KEGG analysis of AMPD1 metabolic pathways

4 Discussion

The present study undertakes a systematic bioinformatics approach to identify functionally important nsSNPs in the human AMPD1 gene to understand how these mutations affect the protein function and structure and hence promote a myoadenylate deaminase deficiency.

A total number of 72 nsSNPs of the AMPD1 gene were submitted for analysis by fourteen different bioinformatics tools eight nsSNPs namely (G246S, K320I, R421W, R458C, R458H, R51C, R757L, and R761H) were predicted as deleterious, associated with disease, highly conserved, and decrease effective stability of the AMPD1 protein. According to some studies, most of the diseases causing and cancer-related mutations are found to destabilize the corresponding protein and decreased protein stability causes: an increase in degradation, misfolding, and aggregation of proteins. While increased protein stability was reported to be accompanied by elevated protein levels and protein dysregulation **(Singh et al., 2010, Agoston et al., 2005).** Those nsSNPs which are located in these conserved regions were reported as very damaging to protein as compared to those in non-conserved regions **(Doniger et al., 2008)**. Therefore, any alterations in the AMPD1 protein structure or function would have an impact on many cellular processes and associated pathways which may lead to the onset of diseases.

Missense mutations of nsSNPs were identified in this study revealing that residues located in a domain that is important for the main activity of the protein mutation might disturb this function. Although mutations introduce a more hydrophobic residue, can result in the loss of hydrogen bonds and/or disturb correct folding. Also, loss of charge can cause a loss of interactions with other molecules or residues. Here we present the results of each residue and discuss the conformational variations:

rs34526199(K320I): this mutation resulted in a change of a Lysine into an Isoleucine at position 320 which is located in a domain that is important for the main activity of the protein. The lysine residue is bigger and it was charged positively, while the isoleucine charge is neutral. Instead of that mutation at this position introduces a more hydrophobic residue. A similar mutation of (K320I) on the AMPD1 gene was identified in Indian people studied for evaluation of genetic myopathies, which revealed that this mutation is likely pathogenic and associated with high CK levels **(Chakravorty et al., 2020)**.

rs35859650(R421W): change of an Arginine into a Tryptophan at position 421 in the AMPD1 gene will result in: loss of Arginine charge, a bigger size of Tryptophan residue which might lead to bumps. Also, the mutation introduces a more hydrophobic residue. Based on conservation scores, R421 is highly conserved and exposed, this mutation is probably damaging to the AMPD1.

rs187527797(R458C): the mutation of an Arginine into a Cysteine at position 458 might result in a disturbed function that is important for the main activity of the protein, loss of arginine charge, and more hydrophobic residue introduces. Instead of that cysteine residue has a smaller size than arginine. Based on conservation scores, R458 is a highly conserved and exposed residue. This mutation is probably leading to the loss of activity of AMPD1.

rs121912682(R458H): change of an Arginine into a Histidine at position 458 might result in a disturbed function that is important for the main activity of the protein, losing charge of an arginine residue, and the smaller size of histidine. Based on conservation scores, R458 is a highly conserved and exposed residue. This mutation is probably leading to the loss of activity of AMPD1.

rs369335745(R51C): this mutation resulted in a change of an Arginine into a Cysteine at position 51. Cysteine residue is located in a domain that is important for the main activity of the protein also it is smaller than Arginine which might lead to a loss of interactions. Arginine will lose its charge and more hydrophobic residue will be introduced. As predicted by Mutpred, this substitution resulted in a loss of acetylation and methylation at K56. This mutation is probably leading to the loss of activity of AMPD1.

rs146036570(R757L): a mutation of an Arginine into a Leucine at position 757 will result in: disrupting the main activity of the protein, loss of Arginine charge, and smaller Leucine molecule; which can result in loss of interaction with other molecules. Although, this mutation at this position introduces a more hydrophobic residue.

rs142643298(R761H): this mutation resulted in a change of an Arginine into a Histidine at position 761. The mutated residue is located in a domain that is important for the main activity of the protein. There may also be other changes such as loss of the

arginine charge and a change in the size of the
It has been predicted that the AMPD1 protein has strong interactions with many proteins involved in the purine nucleotide biosynthesis process, including ADSS, ADSSL1, PURH, and HPRT1. Though Adenine phosphoribosyl transferase (APRT) catalyzes a salvage reaction that generates AMP. Hence, the purine nucleotide cycle is central to skeletal muscle energy production **(Bhagavan and Ha, 2015)**.

When a mutation is predicted to alter disorders in the protein structure, it is considered likely to cause disease. Five of the most common mutations (R→W, R→C, E→K, R→H, R→Q) collectively account for 44% of all deleterious disorder-to-order transitions **(Vacic et al., 2012)**. Altered protein disorder or ordered interface and loss of intrinsic disorder were observed for substitutions (K320I, R421W, R458C, R458H, R757L, and R761H).

Post-translational modifications of proteins, such as protein methylation, play an important role in the epigenetic control of gene expression, genome stability, signaling pathways, and regulation of protein–protein interactions. Several studies have reported that methylated proteins are involved in human diseases like cancer **(Poulard et al., 2016, Han et al., 2019)**. Loss of methylation was observed for substitution (R51C).

Gains relative solvent accessibility as it becomes exposed from buried, making it more possible to have active site activity **(Malleshappa Gowder et al., 2014)**. This molecular mechanism was observed for (R421W) substitution.

Finally, Toyama et.al, have discovered two novel non-synonymous variations in the human AMPD1 gene, including K287I and M310I in addition to the known variants C34T, C143T, and C235T **(Toyama et al., 2004)**. There were other nsSNPs detected in this study, possibly as a result of different methods of analysis used to require the SNPs. A limitation of the current study is that no molecular studies have been conducted to confirm that this mutation causes myoadenylate deaminase deficiency. Despite this, the findings of this study will stimulate further research to detect this mutation in myoadenylate deficiency patients.

5. Conclusions

The present study undertakes a systematic bioinformatics approach to identify functionally important nsSNPs in the human AMPD1 gene to understand how these mutations affect the protein function and structure and hence promote a myoadenylate deaminase deficiency.

6. Recommendations

A genome-wide association study with ascertaining phenotype as well as muscle imaging and histidine residue.
functional studies are necessary to verify this group of nsSNPs that can be associated with a myoadenylate deficiency or other disorders.

Abbreviation

AMPD1	Adenosine monophosphate deaminase **1**
ATP	Adenosine triphosphate
DNA	Deoxyribonucleic acid
IMP	Inosine monophosphate
MAD	Myoadenylate deaminase deficiency
MAF	Minor allele frequency
nsSNPs	non-synonymous single nucleotide polymorphism.
RNA	Ribonucleic acid

References

ABE, M., HIGUCHI, I., MORISAKI, H., MORISAKI, T. & OSAME, M. 2000. Myoadenylate deaminase deficiency with progressive muscle weakness and atrophy caused by new missense mutations in AMPD1 gene: case report in a Japanese patient. *Neuromuscul Disord,* 10**,** 472-7.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nat Methods,* 7**,** 248-9.

AGOSTON, A. T., ARGANI, P., YEGNASUBRAMANIAN, S., DE MARZO, A. M., ANSARI-LARI, M. A., HICKS, J. L., DAVIDSON, N. E. & NELSON, W. G. 2005. Increased protein stability causes DNA methyltransferase 1 dysregulation in breast cancer. *J Biol Chem,* 280**,** 18302-10.

AHMETOV, II & FEDOTOVSKAYA, O. N. 2015. Current Progress in Sports Genomics. *Adv Clin Chem,* 70**,** 247-314.

ASHKENAZY, H., EREZ, E., MARTZ, E., PUPKO, T. & BEN-TAL, N. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res,* 38**,** W529-33.

BHAGAVAN, N. V. & HA, C.-E. 2015. Chapter 25 - Nucleotide Metabolism. *In:* BHAGAVAN, N. V. & HA, C.-E. (eds.) *Essentials of Medical Biochemistry (Second Edition).* San Diego: Academic Press.

CAPRIOTTI, E., CALABRESE, R., FARISELLI, P., MARTELLI, P. L., ALTMAN, R. B. & CASADIO, R. 2013. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics,* 14 Suppl 3**,** S6.

CAPRIOTTI, E., FARISELLI, P. & CASADIO, R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res,* 33**,** W306-10.

CHAKRAVORTY, S., NALLAMILLI, B. R. R., KHADILKAR, S. V., SINGLA, M. B., BHUTADA, A., DASTUR, R., GAITONDE, P. S., RUFIBACH, L. E., GLOSTER, L. & HEGDE, M. 2020. Clinical and Genomic Evaluation of 207 Genetic Myopathies in the Indian Subcontinent. *Front Neurol,* 11**,** 559327.

CHENG, J., MORISAKI, H., SUGIMOTO, N., DOHI, A., SHINTANI, T., KIMURA, E., TOYAMA, K., IKAWA, M., OKABE, M., HIGUCHI, I., MATSUO, S., KAWAI, Y., HISATOME, I., SUGAMA, T., HOLMES, E. W. & MORISAKI, T. 2014. Effect of isolated AMP deaminase deficiency on skeletal muscle function. *Mol Genet Metab Rep,* 1**,** 51-59.

CHOI, Y. & CHAN, A. P. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics,* 31**,** 2745-7.

CHU, D. & WEI, L. 2019. Nonsynonymous, synonymous, and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. *BMC Cancer,* 19**,** 359.

COLEY, W., RAYAVARAPU, S., PANDEY, G. S., SABINA, R. L., VAN DER MEULEN, J. H., AMPONG, B., WORTMANN, R. L., RAWAT, R. & NAGARAJU, K. 2012. The molecular basis of skeletal muscle weakness in a mouse model of inflammatory myopathy. *Arthritis Rheum,* 64**,** 3750-9.

DONIGER, S. W., KIM, H. S., SWAIN, D., CORCUERA, D., WILLIAMS, M., YANG, S. P. & FAY, J. C. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet,* 4**,** e1000183.

FENG, A.-F., LIU, Z.-H., ZHOU, S.-L., ZHAO, S.-Y., ZHU, Y.-X. & WANG, H.-X. 2017. Effects of AMPD1 gene C34T polymorphism on cardiac index, blood pressure, and prognosis in patients with cardiovascular diseases: a meta-analysis. *BMC cardiovascular disorders,* 17**,** 174-174.

FERRER-COSTA, C., GELPÍ, J. L., ZAMAKOLA, L., PARRAGA, I., DE LA CRUZ, X. & OROZCO, M. 2005. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics,* 21**,** 3176-3178.

HAN, D., HUANG, M., WANG, T., LI, Z., CHEN, Y., LIU, C., LEI, Z. & CHU, X. 2019. Lysine methylation of transcription factors in cancer. *Cell Death Dis,* 10**,** 290.

HECHT, L., WASS, J., KELLY, L., CLEVENGER-FIRLEY, E. & DUNN, C. 2013. SNAP-Ed Steps to Health inspires third graders to eat smart and move more. *Journal of Nutrition Education and Behavior,* 45**,** 800-802.

HOOD, L. & ROWEN, L. 2013. The Human Genome Project: big science transforms biology and medicine. *Genome Med,* 5**,** 79.

KANEHISA, M., FURUMICHI, M., TANABE, M., SATO, Y. & MORISHIMA, K. 2017. KEGG: new perspectives on genomes, pathways, diseases, and drugs. *Nucleic Acids Res,* 45**,** D353-d361.

LI, B., KRISHNAN, V. G., MORT, M. E., XIN, F., KAMATI, K. K., COOPER, D. N., MOONEY, S. D. & RADIVOJAC, P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics,* 25**,** 2744-50.

LIM, L., PALAYER, M., BRUNEAU, A., LETOURNEL, F., LE MARÉCHAL, C., SIMARD, G., REYNIER, P., HOMEDAN, C. & NADAJ-PAKLEZA, A. 2017. Myoadenylate deaminase deficiency: a frequent cause of muscle pain A case detected by exercise testing. *Ann Biol Clin (Paris),* 75**,** 445-449.

MALLESHAPPA GOWDER, S., CHATTERJEE, J., CHAUDHURI, T. & PAUL, K. 2014. Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. *ScientificWorldJournal,* 2014**,** 971258.

MI, H., POUDEL, S., MURUGANUJAN, A., CASAGRANDE, J. T. & THOMAS, P. D. 2016. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res,* 44**,** D336-42.

NG, P. C. & HENIKOFF, S. 2001. Predicting deleterious amino acid substitutions. *Genome Res,* 11**,** 863-74.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem,* 25**,** 1605-12.

POULARD, C., CORBO, L. & LE ROMANCER, M. 2016. Protein arginine methylation/demethylation and cancer. *Oncotarget,* 7**,** 67532-67550.

ROBERT, F. & PELLETIER, J. 2018. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. *Front Genet,* 9**,** 507.

SABINA, R. L., MORISAKI, T., CLARKE, P., EDDY, R., SHOWS, T. B., MORTON, C. C. & HOLMES, E. W. 1990. Characterization of the human and rat myoadenylate deaminase genes. *J Biol Chem,* 265**,** 9423-33.

SINGH, S. M., KONGARI, N., CABELLO-VILLEGAS, J. & MALLELA, K. M. 2010. Missense mutations in dystrophin that trigger muscular dystrophy decrease protein stability and lead to cross-beta aggregates. *Proc Natl Acad Sci U S A,* 107**,** 15069-74.

TOYAMA, K., MORISAKI, H., KITAMURA, Y., GROSS, M., TAMURA, T., NAKAHORI, Y., VANCE, J. M., SPEER, M., KAMATANI, N. & MORISAKI, T. 2004. Haplotype analysis of human AMPD1 gene: origin of common mutant allele. *J Med Genet,* 41**,** e74.

VACIC, V., MARKWICK, P. R., OLDFIELD, C. J., ZHAO, X., HAYNES, C., UVERSKY, V. N. & IAKOUCHEVA, L. M. 2012. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput Biol,* 8**,** e1002709.

VENSELAAR, H., TE BEEK, T. A., KUIPERS, R. K., HEKKELMAN, M. L. & VRIEND, G. 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics,* 11**,** 548.

VON MERING, C., JENSEN, L. J., SNEL, B., HOOPER, S. D., KRUPP, M., FOGLIERINI, M., JOUFFRE, N., HUYNEN, M. A. & BORK, P. 2005. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res,* 33**,** D433-7.

WANI., M. Y., GANIE, N., MIR, M., DAR, K., RANI, S., MEHRAJ, S., BAQUAL, M. & SAHAF, K. 2018. Advances and applications of Bioinformatics in various fields of life. 5**,** 3-10.