



# BioScientific Review (BSR)

Volume 2 Issue 1, 2020

ISSN(P): 2663-4198 ISSN(E): 2663-4201

Journal DOI: <https://doi.org/10.32350/BSR>

Issue DOI: <https://doi.org/10.32350/BSR.0201>

Homepage: <https://journals.umt.edu.pk/index.php/BSR>

Journal QR Code:



Article:

## *In Silico* Analysis of Variants of Uncertain Significance in *AP4S1* Gene

Author(s):

Sobia Nazir Chaudry  
Ayman Naeem

Ammara Akhtar  
Mureed Hussain

Article DOI:

<https://doi.org/10.32350/BSR.0201.01>

Article QR Code:



Sobia Nazir Chaudry

Chaudry SN, Akhtar A, Naeem A, Hussain M. *In Silico* analysis of variants of uncertain significance in *AP4S1* gene. *BioSci Rev.* 2020;2(1): 01–09.

To cite this article:

[Crossref](#)



A publication of the  
Department of Life Sciences, School of Science  
University of Management and Technology, Lahore, Pakistan

# ***In Silico* Analysis of the Variants of Uncertain Significance in AP4SI Gene**

Sobia Nazir Chaudry, Ammara Akhtar, Ayman Naeem, Mureed Hussain\*

Department of Life Sciences, University of Management and Technology, Lahore, Pakistan

\*Corresponding author: [mureed.hussain@umt.edu.pk](mailto:mureed.hussain@umt.edu.pk)

---

## **Abstract**

---

Hereditary spastic paraplegia is a group of heterogeneous neurological disorders with genetic etiologies. It is characterized by spasticity in lower limbs along with neurological complications. Sequencing technologies have identified numerous disease causing variants in *AP4SI* gene. However, many very low frequency variations in *AP4SI* have the potential to cause hereditary spastic paraplegia in a recessive inheritance manner. This study was designed to identify these potential disease causing variants in *AP4SI* gene using *in silico* tools. These tools predict the effects of deleterious variants on protein function and pre-mRNA splicing. To predict the pathogenicity of missense variants PhD-SNP<sup>®</sup>, PROVEAN, SNPs&GO, and CADD were used. Splice site variants were analyzed using Spliceman, SPiCE, and Human Splice Finder (HSF). *In silico* analysis identified six missense and five splice site variants with the potential to cause hereditary spastic paraplegia.

**Keywords:** *AP4SI*, heredity spastic paraplegia, *in silico* prediction, SNPs

---

## **1. Introduction**

Hereditary spastic paraplegia (HSP) is a collection of neurological diseases from a clinical and genetic perspective. It impairs human corticospinal tracts and results in progressive lower limb spasticity and weakness [1]. It is known to be genetically inherited by the offspring and the inheritance pattern can be autosomal recessive, autosomal dominant, X-linked recessive and *de novo*. HSP is categorized into a complicated form (complex) and an uncomplicated form (pure) [2]. Clinical abnormalities associated with it include ataxia, thin corpus callosum, peripheral neuropathy and cognitive dysfunction. The disease is known to be rare across various human populations with a prevalence rate of ~3–9/100,000 [3].

Mutations in more than sixty genes are known to cause HSP [4, 5]. A number of

pathogenic mechanisms are involved in this disease including intracellular active transport, endoplasmic reticulum shaping, lipid synthesis and metabolism, endolysosomal trafficking pathway, autophagy, myelination, motor based transportation, and mitochondrial function [6, 3, 7, 8]. *AP4SI* protein is a part of heterotetrameric AP4 complex and the variants of *AP4SI* are known to cause HSP [9, 10]. AP4 complex is expressed in neuronal cells throughout embryological and postnatal development. It plays a key role in protein processing, sorting, trafficking and vesicle formation.

For the identification of missense variants in *AP4SI* gene, CADD, PhD-SNP<sup>®</sup>, PROVEAN, and SNP&GO tools were used. CADD is used for annotating and interpreting genetic variants in human beings. Its major benefit is that it

objectively provides many annotations for every single variant and transforms them into a single measure PHRED or C-score. CADD also provides genetic architecture, which is beyond the scope of any currently known single annotation method, while supporting GRch37 and GRch38 genome assembly [11]. PhD-SNP<sup>g</sup> is a machine learning method based only on sequenced-based features. It is used for the rapid analysis of single nucleotide polymorphism. It is also used as a benchmark for the development of many other tools. This user friendly interface predicts the effect of single nucleotide variants, both in coding and non-coding regions [12]. PROVEAN web server makes predictions about single amino acid and nucleotide substitution for both human and mouse genomes. It has improved running time in the current version. Protein variation effect analyzer uses a preset cutoff value, that is, 2.5 for high balanced accuracy. It gives an opportunity to its users to set their own cutoff value to attain higher sensitivity and specificity of results [13]. SNPs&GO server uses protein functional annotation to make prediction about damaging and deleterious amino acid substitution. SNPs&GO provides 79% accurate results with sequence-based inputs and 83% accurate results using structure-based inputs [14]. For the identification of splice site variants Spliceman, SPiCE and HSF tools were used. Spliceman is an online tool which predicts how likely a distantly present mutation is to affect splice site in a gene [15]. It provides the results of splicing as a ranked list predicting the effect of point mutation on pre-mRNA. Spliceman accepts input with reference and mutated allele having flanking regions known as hexamers. Higher the L1 distance, higher will be chances of disruption of splice site and vice versa [16]. SPiCE is

another powerful tool for the prediction of splice site variants. SPiCE uses *in silico* predictions from Splice Site Finder and Max Ent Scan based on two threshold values 0.115 and 0.749, respectively [17]. These two threshold values are used to classify variants into the following predictive classes: (i) low: variant with no effect on splicing (ii) medium: variant likely to alter splicing and (iii) high: spliceogenic variant [18, 19]. Human splice finder (HSF) is a diversified predicting tool designed to check the impact of different types of mutations including missense, non-sense and splice site. HSF identifies splicing motifs in any human sequence and also predicts the effects of variants at splice site. It is updated regularly and has proved to be very helpful in research, diagnostic, and therapeutic purposes as well as for Human Variome Project. It gives detailed predicting information about 5'ss, 3'ss, base pair sequences as well as ESE and ESS in a genome [20].

## 2. Material and Methods

### 2.1. Variant Identification

ExAC, gnomAD, dbSNPs and Variation Viewer databases were used to enlist the variants of uncertain significance. gnomAD uses v2 (GRCH37/hg19) which contains largely non-overlapping samples. dbSNPs and Variation Viewer databases contain the nucleotide variants reported in human disorders. The focus of this study was to identify missense and splice site variants. So, all other variants in *AP4SI* gene were excluded. These variants were filtered on the basis of quality and allelic frequency. An allele frequency <0.002 was set for the inclusion of variants and they were further analyzed through CADD.

## 2.2. CADD

For the evaluation of the deleteriousness of SNPs, CADD tool was used. CADD utilizes a wide range of functional categories and is capable of prioritizing functional, deleterious and disease causing variants affecting size and genetic architectures. Variants were uploaded as VCF file and were analyzed following instructions for variant analysis. The results were interpreted from the output file which sorts deleterious variants based on a C-Score [21]. C-Score >20 was set to filter out pathogenic variants.

## 2.3. Protein Stability Analysis

Single amino acid alteration can lead to the loss of protein function. iStable and I-Mutant are protein stability predicting tools. It is necessary to study changes in protein stability via single amino acid substitution that leads to the loss of protein function. The difference in folding between wild type and mutant free energy change is the major contributor towards protein instability. Here, computational tools assist researchers to find the change in protein stability caused by single amino acid alteration without conducting any experimental studies. i-Stable and I-Mutant predict such change using information sequence and structural information regarding protein [22]. iStable combines I-Mutant, MUpro, i-Stable tools and interprets results in the form of increase or decrease in stability [23]. Only those variants were selected which showed a decrease in protein stability.

## 2.4. Splice Site Variants

Splice site variants for *AP4SI* were identified and analyzed using Spliceman, HSF, and SPiCE [24].

## 3. Results

### 3.1. Missense Variant Analysis for *AP4SI* Gene

*AP4SI* variants were obtained from ExAC, gnomAD and dbSNPs and a total of 374 missense variants were selected for computational analysis. After applying the filter of allelic frequency <0.002, 12 variants were excluded and the remaining 362 variants were analyzed with *in silico* tools. Variants were annotated with CADD and it selected the variants with missense consequence and C-Score >15. It resulted in the exclusion of 262 variants and the remaining 100 variants were further analyzed and filtered on the basis of PhD-SNP<sup>g</sup> score >0.5, PROVEAN score <-2.5, and SNPGO score >0.5. As a result, the six most pathogenic missense variants were obtained. Details of these six pathogenic variants are provided in Table 1. I-Mutant and i-Stable tools are used to study the effects of these missense variants on protein stability (Table 3).

### 3.2. Splice Site Variants Analysis of *AP4SI* Gene

A total of 420 splice site variants of *AP4SI* gene were identified through ExAC, gnomAD and dbSNPs. Variants with allele frequency <0.002 were excluded and that left us with 408 variants. CADD was used to analyze these remaining variants. Among these variants, 471 showed a C-Score above 20 and only 5 were present in the canonical splice site. The effect of impaired splicing due to these variants was further analyzed using Spliceman, HSF and SPiCE. The results of these variants are depicted in Table 2.

**Table 1.** List of Six Missense Variants in *AP4SI* Predicted as Deleterious

Chr	Pos	Ref	Alt	Amino Acid Change	PhD-SNPg Prediction (Score)	PROVEAN Prediction (Score)	SNPs&GO Prediction (Score)	CADD C-Score	gnomAD Allele Frequency (10 <sup>-5</sup> )
14	31539089	G	A	p.Arg60Gln	Pathogenic (0.992)	Deleterious (-3.97)	effect (79)	33	1.06
14	31535517	T	C	p.Cys39Arg	Pathogenic (0.989)	Deleterious (-11.64)	effect (67)	28.2	1.19
14	31535494	T	C	p.Leu31Pro	Pathogenic (0.986)	Deleterious (-5.8)	effect (61)	28.5	0.398
14	31539088	C	T	p.Arg60Trp	Pathogenic (0.983)	Deleterious (-7.98)	effect (93)	33	1.99
14	31542114	G	A	p.Glu77Lys	Pathogenic (0.947)	Deleterious (-3.71)	effect (84)	31	3.18
14	31535436	G	A	p.Gly12Arg	Pathogenic (0.992)	Deleterious (-7.97)	effect (52)	29.5	6.37

**Table 2.** List of Five Predicted Deleterious Splice Site Variants in *AP4SI*

Splice Site Change	Spliceman Prediction		SPiCE Prediction				HSF Prediction		
	L1 Distance	Ranking L1	SSF_Wt	SSF_Mut	MES_Wt	MES_Mut	Probability	Wt_Score	% Variation
c.138+1G>A	36747	76%	87.73	0	10.08	1.9	High	89.42	-30
c.138+2T>G*	35374	69%	87.73	0	10.08	2.43	High	89.42	-2.16



Splice Site Change	Spliceman Prediction		SPiCE Prediction				HSF Prediction		
	L1 Distance	Ranking L1	SSF_Wt	SSF_Mut	MES_Wt	MES_Mut	Probability	Wt_Score	% Variation
c.139-2A>G	35069	68%	78.49	0	7.56	-0.39	High	79.25	-36.53
c.225+1G>A	32619	55%	91.29	0	10.06	1.88	High	95.9	-27.98
c.294+1G>T*	32634	55%	88.3	0	10.36	1.85	High	93.49	-28.71

\*; variants reported in ClinVar

**Table 3.** Stability Analysis of Six Missense Variants in *AP4S1*

Amino Acid Change	i-Mutant	DDG	Mupro	Consurf Score	IStable	Consurf Score
p.Arg60Gln	Decrease	-0.67	Decrease	-0.682	Decrease	0.7865
p.Cys39Arg	Decrease	-0.31	Decrease	-0.031	Decrease	0.7613
p.Leu31Pro	Decrease	-1.68	Decrease	-1	Decrease	0.805
p.Arg60Trp	Decrease	-0.39	Decrease	-0.06	Decrease	0.7748
p.Glu77Lys	Decrease	-0.57	Decrease	-1	Decrease	0.8107
p.Gly12Arg	Decrease	-0.46	Decrease	-0.535	Decrease	0.8529

#### 4. Discussion

Bioinformatics plays a vital role not only in the analysis of modern sequencing data but also in data interpretation by providing meaningful results. Currently, *in silico* approaches are used for predicting whether a given SNP is associated with a particular disorder or not, while considering different parameters such as clinical, demographic, structural, and bioinformatic during the analysis of the results. [25]. *In silico* studies focus on the functional approach to identify deleterious mutations. Special concern should be given to single nucleotide polymorphism that occurs in coding and non-coding regions causing alteration in RNA which should be verified through functional studies. In this study, *in silico* was utilized to determine the deleterious effects of missense and splice site variations with uncertain significance in *AP4SI* gene.

Adaptor proteins including adaptor protein-4 are believed to be new players in endosomal trafficking. The role of AP1-3 complexes was studied in detail. AP-4 complex plays a specific role in endosome membrane trafficking (from trans Golgi network to endosome) of particular proteins, such as amyloid precursor protein (APP) has a role in basolateral sorting in polarized cells as well as in vesicle formation and the selective protein's uptake into the vesicles [9]. Six missense and five splice site variants in *AP4SI* (NM\_007077) were reported in this study. Two splice site variants found in the study have been reported previously in ClinVar [26] database and cause HSP. It shows the authenticity of the methodology followed in this study. Three more splicing variants in *AP4SI* gene were reported in the current study as novel

variants having no record in ClinVar database.

In our research, we identified nonsynonymous and splicing variants in *AP4SI* that can be deleterious in a homozygous form and cause HSP. For the selection of variants, a stringent criterion was adopted for each predictive tool and highly deleterious missense and splice site variants of *AP4SI* gene were obtained. To check the impact of missense variations on protein stability, iStable and I-Mutant tools were used.

#### 5. Conclusion

In this study, variants caused by amino acid substitution and canonical nucleotide changes were analyzed using authentic computational tools. All identified variants have the potential to cause an autosomal recessive form of hereditary spastic paraplegia. Finally, these mutations were compared with data in ClinVar database to check whether these mutations have been reported previously or not. *In vivo* analysis can further confirm the authenticity of this study. All reported variants are rare with allelic frequency <0.002 in gnomAD and a C-Score greater than 20, indicating these variants as deleterious. The current study highlights the importance of computational modelling to identify deleterious variants in recessive disorders.

#### 6. Conflict of Interest

None to declare.

#### References

- [1] Salinas S, Proukakis C, Crosby A, Warner TT. Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms. *Lancet Neurol.* 2008;7(12): 1127-1138.

- [2] Moreno-De-Luca A, Helmers SL, Mao H, et al. 2011. Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J Med Genet.* 2011;48(2): 141-144.
- [3] Blackstone C. Cellular pathways of hereditary spastic paraplegia. *Annu Rev Neurosci.* 2012;35: 25-47.
- [4] Morais S, Raymond L, Mairey M, et al. Massive sequencing of 70 genes reveals a myriad of missing genes or mechanisms to be uncovered in hereditary spastic paraplegias. *Eur J Hum Genet.* 2017;25(11): 1217-1228.
- [5] Boutry M, Morais S, Stevanin G. Update on the genetics of spastic paraplegias. *Curr Neurol Neurosci Rep.* 2019;19(4): 18.
- [6] Vitner EB, Platt FM, Futerman AH. Common and uncommon pathogenic cascades in lysosomal storage diseases. *J Biol Chem.* 2010;285(27): 20423-20427.
- [7] Yalçın B, Zhao L, Stofanko M, et al. Modeling of axonal endoplasmic reticulum network by spastic paraplegia proteins. *Elife.* 2017;6: e23882.
- [8] Davies AK, Itzhak DN, Edgar JR, et al. AP-4 vesicles contribute to spatial control of autophagy via RUSC-dependent peripheral delivery of ATG9A. *Nat Commun.* 2018;9(1): 1-21.
- [9] Hirst J, Irving C, Borner GH. Adaptor protein complexes AP-4 and AP-5: new players in endosomal trafficking and progressive spastic paraplegia. *Traffic.* 2013;14(2): 153-164.
- [10] Abou Jamra R, Philippe O, Raas A-R, et al. Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am J Hum Genet.* 2011; 88(6): 788-795.
- [11] 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): 310-315.
- [12] Capriotti E, Fariselli P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res.* 2017;45(W1): W247-W252.
- [13] Yongwook C, Agnes PC. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 2015; 31(16): 2745–2747.  
<https://doi.org/10.1093/bioinformatics/btv195>
- [14] Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs & GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics.* 2013; 14(3): S6.
- [15] Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet.* 2018; 59(3): 253-268.
- [16] Lim KH, Fairbrother WG. Spliceman - a computational web server that predicts sequence



- variations in pre-mRNA splicing. *Bioinformatics*. 2012; 28(7): 1031-1032.
- [17] Leman R, Gaildrat P, Le Gac G, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico / in vitro studies: an international collaborative effort. *Nucleic Acids Res*. 2018; 46(15): 7913-7923.
- [18] Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*. 1987; 15(17): 7155-7174.
- [19] Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004; 11(2-3): 377-394.
- [20] Desmet F-O, Hamroun D, Lalande M, Collod G-B, Claustres M, Bérout C. Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009; 37(9): 67-e67.
- [21] Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2018; 47(D1): D886-D894.
- [22] Chen CW, Lin MH, Liao CC, Chang HP, Chu YW. iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules. *Comput Struct Biotechnol J*. 2020; 18: 622-630.
- [23] Capriotti E, Fariselli P, Casadio R. I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*. 2005; 33: W306-W310.
- [24] van den Berg BA, Reinders MJ, Roubos JA, de Ridder D. SPiCE: a web-based tool for sequence-based protein classification and exploration. *BMC Bioinf*. 2014; 15(1): 93.
- [25] Tchernitchko D, Goossens M, Wajcman H. In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics. *Clin Chem*. 2004; 50(11): 1974-1978.
- [26] Landrum MJ, Kattman BL. Clin Var at five years: delivering on the promise. *Hum Mutat*. 2018; 39(11): 1623-1630.