EAS Journal of Biotechnology and Genetics
Abbreviated Key Title: EAS J Biotechnol Genet
ISSN: 2663-189X (Print) & ISSN: 2663-7286 (Online)
Published By East African Scholars Publisher, Kenya

Volume-2 | Issue-3|May-June-2020 |

OPEN ACCESS

Research Article

Analysis of Single Nucleotide Polymorphisms in Human Alanyl-tRNA Synthetase (*AARS)* Gene: an in Silico Approach

Abdulrhman A. Elshiekh[1], Afra M. Albkrye[1], Mona Abdelrahman M.Khaeir[1] and Hind A.Elnasri*[1]

[1]Department of Molecular Biology and Bioinformatics, College of Veterinary Medicine, University of Bahri- Khartoum – Sudan

Article History
Received: 04.05.2020
Accepted: 28.05.2020
Published: 09.06.2020

Journal homepage:
https://www.easpublisher.com/easjbg

Quick Response Code

Abstract: Charcot-Marie-Tooth (CMT) disease is an autosomal recessive disease that manifests at the age 15 years or late in life. It is characterized by muscle weakness and atrophy of the peripheries, and damages to the peripheral nervous system. It results from mutations in the *AARS* gene which causes neuro degeneration. The aim of this study was to determine the effect of deleterious SNPs (Single Nucleotide Polymorphisms) in the coding region of the Human *AARS* gene on protein structure and function using computational analysis. The nonsynonymous SNPs in the gene were downloaded from the dbSNP database in 2019.These SNPs were analyzed using various software including: Gene MANIA which provides information about the proteins interaction, Sorting Tolerant From Intolerant (SIFT),Polymorphism PhenotypingV2(Polyphen-2**),**Protein Variation Effect Analyzer v1.1 (PROVEAN), SNPs & GO, and Phd-SNP (Predictor of Human Deleterious Single Nucleotide Polymorphism)to sort out deleterious form non-deleterious SNPs. Then the effect of SNPs on the stability of the alanyl-aminoacyl tRNA synthetase enzyme were predicted using I-Mutant3.0 and Mupro, and lastly project HOPE was used to predict the effect of SNPs on the structure of the enzyme. The total number of the SNPs obtained was 11849, only 15 SNPs were found to be disease related using various software used. Using the stability software, 14 SNPs were found to decrease the protein stability. A total of 13 SNPs were predicted to be disease related and have not been reported before.

Keywords: *AARS* gene, CMT (Charcot-Marie-Tooth), In-silico analysis, SNPs, Translational Bioinformatics.

INTRODUCTION:

Charcot–Marie–Tooth disease (CMT) is the most common inherited neuropathy with heterogeneous clinical presentation and genetic background (Sindhu R , 2017). It is characterized by distal symmetric polyneuropathy (DSP), muscle weakness and wasting, diminished deep tendon reflexes, impaired sensation in the extremities(Heather M *et al* ., ,2013).The estimated frequency of the disease is 1/2500 (Philippe L *et al* ., 2010). CMT has three main subtypes: CMT1 demyelinating polyneuropathy which is based on slow conduction velocity of 38 meter per second in the upper extremities, CMT2 axonal polyneuropathy with reduced sensory nerve or compound muscle action potential (CAMP) amplitude indicating axonal loss and CMT3 which includes recessive disorders with a severe phenotype(Sindhu R, 2017) .

AARS (Alanyl-tRNA Synthetase) gene is located in chromosome 16q22, and is composed of 21 exons and codes for a protein made up from 968aminoacids(Philippe L *et al.,* 2010). Aminoacyl-tRNA Synthetase plays a central role in the process of protein synthesis, since it is responsible for catalyzing the attachment of the correct amino acid to its cognate tRNA through an esterification reaction at the 3' end of tRNA(Yaxue Z *et al* ., 2014). Up until now the relationship between these ubiquitously expressed enzymes and CMT are not known.

Single nucleotide polymorphism (SNPs) are the most common form of genetic variation in humans (about 1:1000 of the average genome). Studies showed that most SNPs are found in 95% of the non-coding region of the genome (HagmannM,1999; Sean M,2005). These SNPs can alter the function of DNA, RNA or protein according to their specific location within the gene. The non-synonymous SNPs can change the amino acid sequence within a protein either by substitution of the amino acid or producing a nonsense mutation(Sean M,2005) . Other SNPs can affect the expression of the gene by interrupting a regulatory region or interfering with the normal splicing of mRNA(Sean M,2005). Recently using translational bioinformatics makes the search for association between disease SNPs and genes easier, especially with the huge amount of data that is obtained from the high through-put technologies being generated every day.

***Corresponding Author:** Hind A.Elnasri

In this study the analysis of non synonymous SNPs in the coding region of the Human *AARS* gene was performed using computational analysis. Also, the effect of the SNPs on the protein structure and stability was investigated.

MATERIALS AND METHODS:

SNPs Retrieval from Database:

All the SNPs were obtained from dbSNP NCBI database (https://www.ncbi.nlm.nih.gov/projects/SNP), during the year 2019. Identification of the damaging nsSNPs was done using several bioinformatics software including:

1. GeneMANIA:

Gene MANIA(http://www.genemania.org)is a flexible, user friendly web interface for generating hypotheses about gene function, analyzing gene lists and prioritizing genes for functional assays. Gene MANIA finds related genes to input genes, using a very large set of functional association data. The input is the gene's name, and the results shows the physical interaction, functional data, and other information related to the specific gene (Warde-Farley D *et al* ., 2010)

2. Functional effect of the SNPs:

2.1. Sorting Intolerant From Tolerant (SIFT): https://sift.bii.a-star.edu.sg/).

SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of each amino acid, using the genome tools in the website. A list of nonsynonymous SNPs (rsID)which were previously obtained from dbSNP database were submitted, and SIFT predicts whether the SNPs are either Deleterious or Tolerated with cutoff point of 0.05.(Jaroslav B,2014). Only the deleterious SNPs were subjected for further analysis.

2.2 Polymorphism Phenotyping V2 (Polyphen-2): (http://genetics.bwh.harvard.edu/pph2/)

Polyphen-2 predicts the possible impact of amino acid substitution on the stability and function of human proteins using structural and comparative evolutionary considerations. The protein sequence, position of the wild type and mutant amino acids were submitted to the software.Polypeh-2 predicts if the SNP may be benign or damaging with a score ranging from 0 to 1 respectively, 0.5 is the cutoff value for discriminating between them (Ivan A *et al* ., 2010)

2.3 Protein Variation Effect Analyzer v1.1(PROVEAN): (http://provean.jcvi.org/index.php)

PROVEAN provides a generalized approach to predict the functional effects of protein sequence variations including single or multiple amino acid substitutions, insertions or deletions. A list containing the protein ID number, the position of the mutation, the wild and the mutant amino acid were submitted , and then PROVEAN predicts whether the variant is deleterious or neutral giving a score and cutoff value of 2.5 (Yongwook C. and Agnes P, 2015).

2.4 SNPs&GO: (http://snps.biofold.org/snps-and-go/snps-and-go.html/)

It is a method that starting from a protein sequence, can predict whether a mutation is disease related or not by exploiting the protein functional annotation, and functionas encoded in the Gene Ontology terms. The protein sequence and each mutation was entered indicating the wild type, the position of mutation and the mutant amino acid, the results will indicate either the mutation is disease or neutral with a probability score of the prediction(Remo C *et al* ., 2009).

2.5 PHD-SNP (Predictor of human deleterious single nucleotide polymorphism): (http://snps.biofold.org/phd-snp/phd-snp.html/).

It is an online tool that uses support vector machine (SVM) method to predict disease associated SNPs using sequence information. The sequence of the protein and the SNPs were the input, and the output reveals if the SNP is disease causing according to the threshold of 0.5. A higher score indicated that the SNP is associated with the disease (Emidio C *et al* .,2013)

3. Effect of SNPs on protein stability:

To study the effect of the SNPs on the stability of the protein two software were used**:**

3.1 I Mutant Suite: **(**http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0 /I-Mutant3.0.cgi/).

I-Mutant is a suite of Support Vector Machine based process. It predicts protein stability changes upon single-site mutations. It also predicts the deleteriousness of the mutation. The input is the protein sequence, the position of the SNP in the protein and the new residue, it predicts that either the mutation increases or decreases the stability of the protein (Emidio C. *et al* ., ,2005).

3.2 MUpro: (http://mupro.proteomics.ics.uci.edu/)

It is a set of machine learning programs to predict how single-site amino acid mutation affects protein stability using Support Vector Machines and Neural Networks. The input was the protein sequence, the wild- and mutant amino acid and their corresponding positions in the sequence, then the output is either an increase or decrease in the stability (Jianlin C *et al* ., 2006).

4. Structural effect of SNPs:

Project HOPE: (http://www.cmbi.ru.nl/hope/)

It is a web server that takes the protein sequence and the mutation of interest as an input and analyzes the mutation using several web servers to get the effect of the mutation on the protein. HOPE predicts

the effect of the mutation on the protein properties depending on the difference between the mutated and wild type amino acid (Hanka V,2015).

RESULTS AND DISCUSSION:

This study revealed that *AARS gene* is associated with 20 other genes, their physical interaction, co-expression and other information are shown in Figure 1, and Table 1.

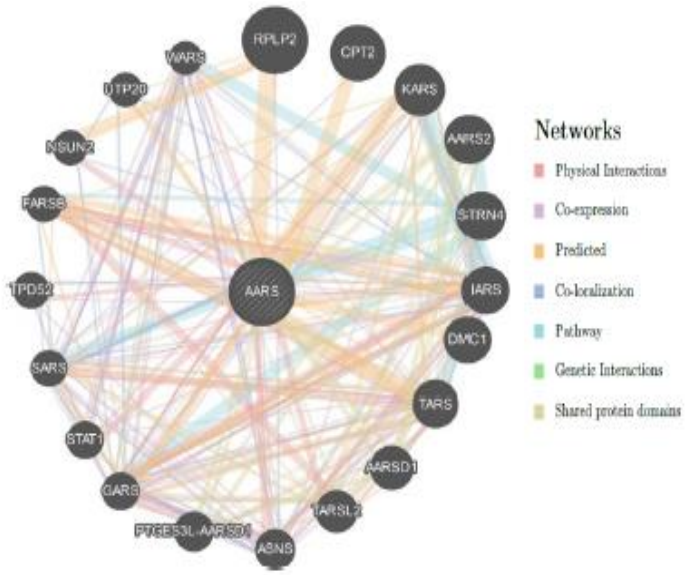

Figure 1: GeneMANIA results of *AARS g*ene.

Table 1: Gene Description Rank using GeneMANIA

Gene	Description
AARS	Alanyl-tRNA synthetase
RPLP2	Ribosomal protein lateral stalk subunit P2
CPT2	Carnitine palmitoyltransferase 2
KARS	Lysyl-tRNA synthetase
AARS2	Alanyl-tRNA synthetase 2, mitochondrial
STRN4	Striatin 4
IARS	Isoleucyl-tRNA synthetase
DMC1	DNA meiotic recombinase 1
TARS	Threonyl-tRNA synthetase
AARSD1	Alanyl-tRNA synthetase domain containing 1
TARSL2	Threonyl-tRNA synthetase-like 2
ASNS	Asparagine synthetase (glutamine-hydrolyzing)
PTGES3LAARSD1	PTGES3L-AARSD1 readthrough
GARS	Glycyl-tRNA Synthetase
STAT1	Signal transducer and activator of transcription 1
SARS	Seryl-tRNA Synthetase
TPD52	Tumor protein D52
FARSB	Phenylalanyl-tRNA synthetase beta subunit
NSUN2	NOP2/Sun RNA methyltransferase family member 2
UTP20	UTP20, small subunit processome component
WARS	tryptophanyl-tRNA synthetase

*AARS2*gene (Alanyl-tRNA Synthetase 2) was found to be associated with this gene .It has the same function as the AARS enzyme but it is located within the mitochondria, it is also linked to another neurodegenerative disorder(Alexandra G *et al* ., ,2011).

dbSNP database showed 11849 SNPs were found in the gene, only 183 were nonsynonymous , 15 where in the 3`UTR region, 3 in the 5`UTR region and 165 in the coding region of the *AARS* gene.

The SNPs in the coding region of the *AARS* gene (a total of 165 SNP) , were analyzed using SIFT and the result showed that only 41 were deleterious. These SNPs were taken for analysis using other software. Polyphen-2 showed that 39 SNPs were damaging and only 2 where benign(Fig.2). On the other hand,

PROVEAN software revealed that 7 SNPs were neutral, and 34 were deleterious (Fig.2). SNPs&GO and PHD-SNP severs showed 25and 29 SNPs were disease relate , respectively but only 24 were predicted to be disease related using both servers. Overall the total number of SNPs predicted to be disease related using all five software were 15 SNPs which are shown in table (2).

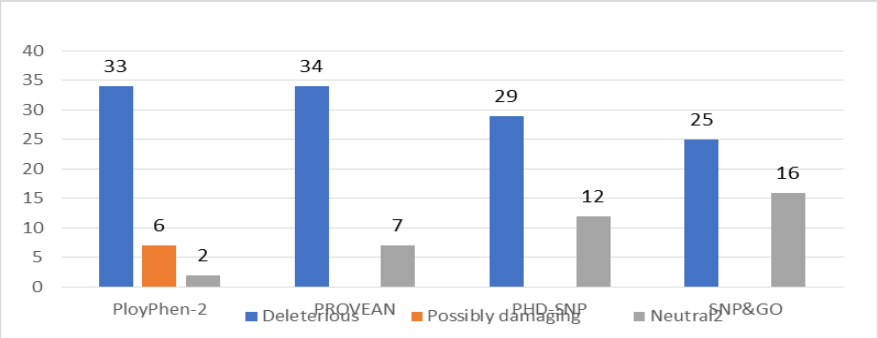

Fig. 2: Result of SNPs analysis using different software.

Table 2: SNPs predicted to be disease causing in all of the software

Amino acid Change	rsID	SIFT	Polyphen2	PROVEAN	Phd-SNP	SNPs&GO
A607T	rs377247120	Deleterious	Possibly Damaging	Deleterious	Disease	Disease
A676V	rs371114066	Deleterious	Probably Damaging	Deleterious	Disease	Disease
G913D	rs369774476	Deleterious	Probably Damaging	Deleterious	Disease	Disease
I52T	rs200582917	Deleterious	Probably Damaging	Deleterious	Disease	Disease
N604S	rs371595630	Deleterious	Probably Damaging	Deleterious	Disease	Disease
N71Y	rs387906792	Deleterious	Probably Damaging	Deleterious	Disease	Disease
R320C	rs138490305	Deleterious	Probably Damaging	Deleterious	Disease	Disease
R329H	rs267606621	Deleterious	Probably Damaging	Deleterious	Disease	Disease
R695Q	rs377378009	Deleterious	Probably Damaging	Deleterious	Disease	Disease
R729W	rs138081804	Deleterious	Probably Damaging	Deleterious	Disease	Disease
R751G	rs143370729	Deleterious	Probably Damaging	Deleterious	Disease	Disease
S731L	rs150873930	Deleterious	Probably Damaging	Deleterious	Disease	Disease
T181M	rs202099051	Deleterious	Probably Damaging	Deleterious	Disease	Disease
Y415C	rs373069396	Deleterious	Probably Damaging	Deleterious	Disease	Disease
Y418C	rs147433234	Deleterious	Probably Damaging	Deleterious	Disease	Disease

Regarding the effect of the SNPs in the protein stability, 14 SNPs (out of the 15 SNPs predicted to be disease related using all software) were found to decrease the activity of the protein using MUpro and I-Mutant3.0 software. Only one SNP(rs202099051)was predicted to increase the stability (Table 3).

Table 3: Stability prediction results

Amino acid Change	rsID	I-Mutant 3.0	Mupro
A607T	rs377247120	Decrease	Decrease
A676V	rs371114066	Decrease	Decrease
G913D	rs369774476	Decrease	Decrease
I52T	rs200582917	Decrease	Decrease
N604S	rs371595630	Decrease	Decrease
N71Y	rs387906792	Decrease	Decrease
R320C	rs138490305	Decrease	Decrease
R329H	rs267606621	Decrease	Decrease
R695Q	rs377378009	Decrease	Decrease
R729W	rs138081804	Decrease	Decrease
R751G	rs143370729	Decrease	Decrease
S731L	rs150873930	Decrease	Decrease
T181M	rs202099051	Increase	Increase
Y415C	rs373069396	Decrease	Decrease
Y418C	rs147433234	Decrease	Decrease

Using project hope software, the results showed that 13 SNPs are in the activity domain of the protein, and in contact with other domains that are involved in binding of the protein. These SNPs can disrupt the interaction between domains and affect signal transduction. Another SNP (rs369774476) only affects activity and function of the protein. While one SNP (rs147433234) is located in the domain and is important for binding of other molecules to the protein.

A previous study reported that there are six families having dominant missense alanyl tRNA synthetase (*AARS*) mutation leading to clinically heterogenous axonal neuropathies (Boglarka B *et al* ., 2015).

Several SNPs in the AARS enzyme were found to be related to CMT, a study involved a French family showed an R329H mutation(Philippe L *et al* ., 2010) which has also been predicted in the current study using the 5 different software. This mutation is found in exon 8, a highly conserved region which is essential for the aminoacylation of tRNAs (Philippe L *et al* ., 2010). The same mutation was also reported among an Australian family (Heather M,2013).

Another SNP (rs387906792) has been previously reported in a Taiwanese family(Kon-Ping L *et al* ., ,2011), which has also been predicted in the current study. This mutation is found in exon 3, which also resides in a highly conserved region, and may alter the protein function. Among the SNPs that have been predicted to be pathogenic in the current study , two SNPs namely (rs267606621) and (rs387906792) were tested in vitro – in a previous study - and showed their ability to reduce the catalytic activity of the enzyme (Heather M, 2013).

Table 4: Project HOPE predictions

SNP ID	Amino acid change	Wild type			Mutant type		
		Size	Charge	Hydrophobicity	Size	Charge	Hydrophobicity
rs377247120	A607T	Smaller	-	More Hydrophobic	Bigger	-	Less Hydrophobic
rs371114066	A676V	Smaller	-	-	Bigger	-	-
rs369774476	G913D	Smaller	Neutral	More Hydrophobic	Bigger	Negative	Less Hydrophobic
rs200582917	I52T	Bigger	-	More Hydrophobic	Smaller	-	Less Hydrophobic
rs371595630	N604S	Bigger	-	Less Hydrophobic	Smaller	-	More Hydrophobic
rs387906792	N71Y	Smaller	-	Less Hydrophobic	Bigger	-	More Hydrophobic
rs138490305	R320C	Bigger	Positive	Less Hydrophobic	Smaller	Neutral	More Hydrophobic
rs267606621	R329H	Bigger	Positive	-	Smaller	Neutral	-
rs377378009	R695Q	Bigger	Positive	-	Smaller	Neutral	-
rs138081804	R729W	Smaller	Positive	Less Hydrophobic	Bigger	Neutral	More Hydrophobic
rs143370729	R751G	Bigger	Positive	Less Hydrophobic	Smaller	Neutral	More Hydrophobic
rs150873930	S731L	Smaller	-	Less Hydrophobic	Bigger	-	More Hydrophobic
rs202099051	T181M	Smaller	-	Less Hydrophobic	Bigger	-	More Hydrophobic
rs373069396	Y415C	Bigger	-	Less Hydrophobic	Smaller	-	More Hydrophobic
rs147433234	Y418C	Bigger	-	Less Hydrophobic	Smaller	-	More Hydrophobic

CONCLUSION

The current study revealed 13 novel nsSNPs related to CMT. Two SNPs R329H and N71Y were previously reported to be pathogenic. Using bioinformatics tool to determine pathogenic SNPs can be a useful tool for diagnosis of disease in the future.

REFERENCES:

1. Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.*, no. SUPPL.76, pp. 1–41.
2. Alexandra G., Henna T., Liliya E., Pekka E., Tuulia H., Tiina O., Riikka H. Hamalainen, Johanna T., Taneli R., Matej O., Riitta K., Outi T., Kalle O.J. Simola, Anders P., Tiina T.,& Anu S. (2011). "Exome Sequencing Identifies Mitochondrial Alanyl-tRNA Synthetase Mutations in Infantile Mitochondrial Cardiomyopathy. *AJHG* pp. 635–642.
3. Boglarka B., Thalia A., Sarah B., Sinead M. Murphy, John M., Michael A., Richard W., Joanna D., David H., Hanns L., Patrick C., & Rita H.(2015). Genotype/phenotype correlations in AARS-related neuropathy in a cohort of patients from the United Kingdom and Ireland. *J. Neurol.*, *262* (8), pp. 1899–1908.
4. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., & Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human mutation*, *30*(8), 1237-1244.
5. Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), 2745-2747.
6. Emidio C., Piero F., & Rita C. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, 33, W306–W310.
7. Emidio C., Russ, B., & Altman, Y. B. (2013). Collective judgment predicts disease-associated single nucleotide variants.*BMC Genomics*, *14* Suppl 3.
8. Hagmann, M.(1999). A good SNP may be hard to find. *Science*. V285:21-22.
9. Heather M. McLaughlin, Reiko S., William G., Thomas E. Wilson, Leslie B., James R. L., Kevin T., Jeffery M. V., Stephan Z., Yi-Chung L., Marina K., Ya-Ming H., & Garth, N. (2013). A Recurrent Loss-of-Function Alanyl-tRNA Synthetase (AARS) Mutation in Patients with Charcot-Marie-Tooth Disease Type 2N (CMT2N). *Human Mutation*. *33* (1), 244–253.
10. Jaroslav, B., Jan, S., Ondrej, S., Antonin, P., Eric, D., Wieben, J. Z., Jan, B., & Jiri D.(2014). PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Comput. Biol.*, *10* (1), pp. 1–11
11. Jianlin C., Arlo R., and Pierre B.,(2006). Prediction of protein stability changes for single-site mutations using support vector machines, *Proteins Struct. Funct. Genet.*, vol. 62, no. 4, pp. 1125–1132.
12. Kon-Ping, L., Bing-Wen, S., Chih-Chao, Y., Li-Wen, H., Ming-Hong, C., I-Hui, L., Antony, A., & Yi-Chung, L. (2011). The mutational spectrum in a cohort of Charcot-Marie-Tooth disease type 2 among the han Chinese in Taiwan, *PLoS One*, *6* (12).
13. Latour, P., Thauvin-Robinet, C., Baudelet-Méry, C., Soichot, P., Cusin, V., Faivre, L., ... & Camu, W. (2010). A major determinant for binding and aminoacylation of tRNAAla in cytoplasmic Alanyl-tRNA synthetase is mutated in dominant axonal Charcot-Marie-Tooth disease. *The American Journal of Human Genetics*, *86*(1), 77-82.
14. Sean, M. (2005). Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. Briefings in Bioinformatics. 6:1, 44-56.
15. Sindhu, R.(2017). "Charcot-Marie-Tooth Disease and Other Genetic Polyneuropathies. Peripheral Nerve and Motor Neuron Disorders. pp. 1360–1377.
16. 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*(1), 548.
17. Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., ... & Maitland, A. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*, *38*(suppl_2), W214-W220.
18. Yaxue, Z., Qingqing, M., Linquan, B., & Huchen, Z. (2014). In Silico discovery of aminoacyl-tRNA synthetase inhibitors. *Int. J. Mol. Sci.*, *15* (1), pp. 1358–1373.