The International Journal of Engineering and Science (IJES)
|| Volume || 9 || Issue || 08 || Series I || Pages || PP 47-51 || 2020 ||
ISSN (e): 2319-1813 ISSN (p): 20-24-1805

A preliminary Analysis of the Impact of Missense Variants of Unknown Significance in the *APOA1* Gene using prediction softwares

Ingryd de Souza Albuquerque[1], Izabella Ferreira da Silva[1], Lígia Silva Pondé Serra[1], Nathan Gurgel de Albuquerque Silva[1], William Farias Porto[2], Sérgio Amorim de Alencar[2]

[1]*Programa de Graduação em Ciências Biológicas,UniversidadeCatólica de Brasília, Brasília-DF, Brazil*
[2]*Programa de Pós-Graduação em CiênciasGenômicas e Biotecnologia, Universidade Católica de Brasília, Brasília-DF, Brazil*
CorrespondingAuthor: S. A. Alencar

---***ABSTRACT***---
The apolipoprotein A1 (APOA1) is a component of high-density lipoprotein (HDL) in plasma, playing aspecific role in lipid metabolism. Defects in this protein have been associated with several diseases, including familial HDL deficiency, familial visceral amyloidosis and increase risk of cardiovascular disorders. Among the possible causes of these defects are single nucleotide alterations within the coding region of the APOA1 gene, which could affect protein function. This study aims to analyze the possible impact of 15 variants of unknown significance identified within the APOA1 gene. A total of 12 in silico SNV impact prediction tools were used to evaluate the impact of these alterations. The predictions suggested that four of these alterations(Pro123Ser, Arg140Leu, Ser191Asn and Ser191Arg)could result in protein disfunction.

KEYWORDS: *APOA1; apolipoprotein A1; VUS; impact; SNV*

Date of Submission: 30-09-2020 Date of Acceptance: 13-10-2020

I. INTRODUCTION

The *APOA1* gene, located on chromosome 11 (11q23-q24), encodes a preproprotein which is proteolytically processed to generate the mature apolipoprotein A1 (APOA1), composed of 396 amino acids (van der Vorst, 2020). The APOA1 protein is a component of high-density lipoprotein (HDL) in plasma, playing a specific role in lipid metabolism (Breslow et al., 1982). HDL is a molecule that transports cholesterol and certain fats called phospholipids through the bloodstream from the body's tissues to the liver. Once in the liver, cholesterol and phospholipids are redistributed to other tissues or removed from the body(van der Vorst, 2020).Therefore, due to its importance in the process of removing excess cholesterol from cells, defects in this gene are associated with HDL deficiencies, including familial HDL deficiency, familial visceral amyloidosis, Tangier disease, and increase risk of cardiovascular disease(Ertek, 2017).

Several studies have shown that APOA1 dysfunction is associated with mutations in the *APOA1* gene (Dastani et al., 2006; Franceschini, Sirtori, Gianfranceschi, & Sirtori, 1981), including mutations that lead to alterations in a single nucleotide in the coding region of *APOA1*(Ladias, Kwiterovich, Smith, Karathanasis, & Antonarakis, 1990; Strobl et al., 1988; Utermann, Pruin, & Steinmetz, 1979). In fact, a total of 16 missense variants are currently described in the Online Mendelian Inheritance in Man (OMIM) database as associated with APOA1 dysfunctions.

SNVs are the most common kind of variation in the human genome. SNVs that occur in a gene's coding region and result in an amino acid substitution at the coded protein corresponding region are known as missense SNVs. In most cases, this variation is neutral or has little effect on the protein function. However, when this variation causes an alteration in the protein's structure, this change could also result in protein function alteration, which could lead to a disease (Ng & Henikoff, 2006; Yates & Sternberg, 2013).

Several approaches have been used to study the impact of missense SNVs on a protein structure; however testing these impacts in a laboratory can be expensive, so analysis by computational tools has become a powerful and inexpensive approach for preliminary analyses (Shen, Deininger, & Zhao, 2006). As a result, a number of *in silico* tools have been developed to predict the effect of missense SNVs, including methods based on sequence homology (Choi, Sims, Murphy, Miller, & Chan, 2012) supervised-learning (Liu, Wang, Sun, & Zhao, 2014), protein-sequence (Dorn, E Silva, Buriol, & Lamb, 2014), and consensus-based(González-Pérez & López-Bigas, 2011).

The impact of many missense SNVs occurring in the APOA1 gene is currently unknown. These variants are classified as Variants of Unknown Significance (VUSs). In the present study, using an *in silico* approach, the impact of all APOA1 known missense SNVs classified as VUS was evaluated by a combination of prediction tools.

II. METHODOLOGY

SNV Data

The apolipoprotein A1 (*APOA1*) gene data was obtained from the National Center for Biotechnology Information (NCBI) (access codes: NP_001304946.1, NP_000030.1 and NP_001304950.1). The information about the missense SNVs was collected from the Leiden Open Variation Database (LOVD v.3.0)(Fokkema et al., 2011), selecting only variations classified as "VUS" in the Clinical classification column.

SNV impact prediction tools

All selected missense SNVs were submitted to twelve protein analysis programs. The programs were split into three categories, each containing four programs: sequence homology-based methods, supervised learning methods and consensus-based methods.

Sequence homology-based methods

Disease-causing missense SNVs tend to occur at evolutionarily conserved positions that have an essential role in the structure and/or function of the encoded protein (Miller & Kumar, 2001). Therefore, information contained in multiple sequence alignments (MSAs) of homologous protein sequences can help in understanding contemporary deleterious variations in humans. A missense SNV can lead to an amino acid with altered physicochemical properties compared to the original one and this change can in turn be used to predict functional consequence to the protein (Ng & Henikoff, 2006). The following algorithms are based on these principles and combine MSAs, generated through a variety of methods, with scoring functions based on measures of amino acid similarity to produce functional predictions: Sorting Intolerant From Tolerant (SIFT) (Kumar, Henikoff, & Ng, 2009), Provean(Choi et al., 2012), MutationAssessor(Reva, Antipin, & Sander, 2011) and Panther (Mi et al., 2005).

Supervised learning methods

The supervised learning methods include Neural Networks: (SNAP) (Bromberg, Yachdav, & Rost, 2008) and Support Vector Machines: (PhD-SNP) (E Capriotti, Calabrese, & Casadio, 2006), SuSPect(Yates, Filippis, Kelley, & Sternberg, 2014)and I-Mutant 2.0 (Emidio Capriotti, Fariselli, & Casadio, 2005). In neural networks and support vector machine methods, two training sets are constructed: one containing variants associated with disease and another without disease association. The conservation patterns and physical-chemical properties of the variants on both sets are assessed and used to program the algorithm to “learn” the difference between the variants in the different sets.

Consensus-based methods

There are currently many computational tools widely employed for the prediction of the effects of mutations on protein function. The following softwares combine a variety of methods into a consensus classifier, resulting into significantly improved prediction performance: Condel(González-Pérez & López-Bigas, 2011), PON-P2 (Niroula, Urolagin, & Vihinen, 2015) SNP-effect (De Baets et al., 2012) and SNP&GO (Emidio Capriotti et al., 2013).

Evolutionary conservation analysis

Consurf(Celniker et al., 2013) is a tool used to estimate the evolutionary conservation of amino acid or nucleic acid positions in their respective molecules based on phylogenetic relationships between homologous sequences. It was used to assess the conservation of amino acid residues in the APOA1 sequence. The following parameters were set: CSI-BLAST algorithm for the homolog search, using 3 iterations with an E-value cutoff of 0.0001, against the UNIREF-90 protein database.

III. RESULTS AND DISCUSSION

A total of 95 unique *APOA1* variants are currently displayed at the LOVD database. From these, 15 SNVs classified as VUSs (Val10Met, Arg19Trp, Pro28Arg, Val43Met, Phe81Tyr, Pro123Ser, Arg140Leu, Leu41Ile, Leu150Phe, His159Arg, Glu54Gly, Met172Val, Ala79Thr, Ser191Asn and Ser191Arg) were retrieved for further analysis.

The 15 SNVs were analyzed using the prediction programs divided into three different approaches (sequence homology-based methods, supervised learning methods and consensus-based methods). After this analysis, 4 out of 15 missense SNVs were considered to be deleterious by at least three programs in each category: Pro123Ser, Arg140Leu, Ser191Asn and Ser191Arg (Table 1).Due to the lack of a three-dimensional structure representing the whole structure of the APOA1 protein, it was not possible to carry predictive analysis using structure-based methods, or more robust analysis, such as molecular dynamics simulations.

Although X-ray crystallography or Magnetic Resonance are powerful tools in determining protein 3D structures, these methods are time-consuming and expensive, and there are limitations in the crystallization process of many proteins. However, the protein sequence of APOA1 is available and could be used by several programs in order to obtain useful information regarding the potential of the selected SNVs to cause impact on APOA1 structure.

		Sequence-based				SLM-basedFATHMM				Consensus-based			
Nucleotide change	Amino acid change	SIFT	Provean	Mutation Assessor	Panther	SNAP	PhD-SNP	SuSPect	I-Mutant 2.0	SNP-effect	PON-P2	SNPs&GO	CONDEL
c.28G>A	p.Val10Met	N	N	D	N	N	N	D	D	N	N	D	D
c.55C>T	Arg19Trp	N	D	D	N	D	N	N	D	D	D	N	N
c.83C>G	Pro28Arg	D	N	D	D	D	N	N	N	D	N	N	D
c.127G>A	Val43Met	N	N	N	D	N	N	N	D	D	N	D	D
c.242T>A	Phe81Tyr	N	N	N	N	D	N	N	N	D	N	D	D
c.367C>T	Pro123Ser	D	D	D	N	D	D	N	D	D	D	N	D
c.419G>T	Arg140Leu	D	D	D	D	D	N	D	D	D	D	N	D
c.121C>A	Leu41Ile	N	N	N	D	D	N	D	D	N	N	N	/N
c.448C>T	Leu150Phe	N	N	N	N	D	N	D	D	N	N	N	N
c.476A>G	His159Arg	N	N	N	N	N	N	N	D	D	N	N	N
c.161A>G	Glu54Gly	N	N	N	N	D	N	N	D	N	N	N	N
c.514A>G	Met172Val	N	N	N	N	N	N	N	N	N	N	N	N
c.235G>A	Ala79Thr	N	N	N	N	N	N	N	D	N	N	N	N
c.572G>A	Ser191Asn	D	N	D	D	D	N	D	D	D	D	N	D
c.573C>A	Ser191Arg	D	D	D	D	D	N	D	D	D	D	N	D

Table 1. Prediction results of *APOA1* converging deleterious missense VUSs analyzed by 12 Bioinformatics tools classified in four different groups. "D" corresponds to the deleterious classification and "N" to neutral.

In order to verify whether the SNVs are located in important positions for protein function, an evolutionary analysis based on multiple sequence alignment was carried out using the Consurf software. The results of this analysis showed that two of the deleterious SNVs considered to be deleterious by at least three programs in each category (Pro123Ser and Arg140Leu) are located in highly conserved regions of the APOA1 protein, with conservation value of 8. The other two (Ser191Asn and Ser191Arg) showed medium level of conservation (value of 5) (Figure 1).

Thus, considering both the predictions carried out by the 12 programs and the evolutionary analysis, our results suggest that three missense SNVs located within the APOA1 protein could impact its function. Hence, these variants could be associated to diseases related to APOA1 disfunction.

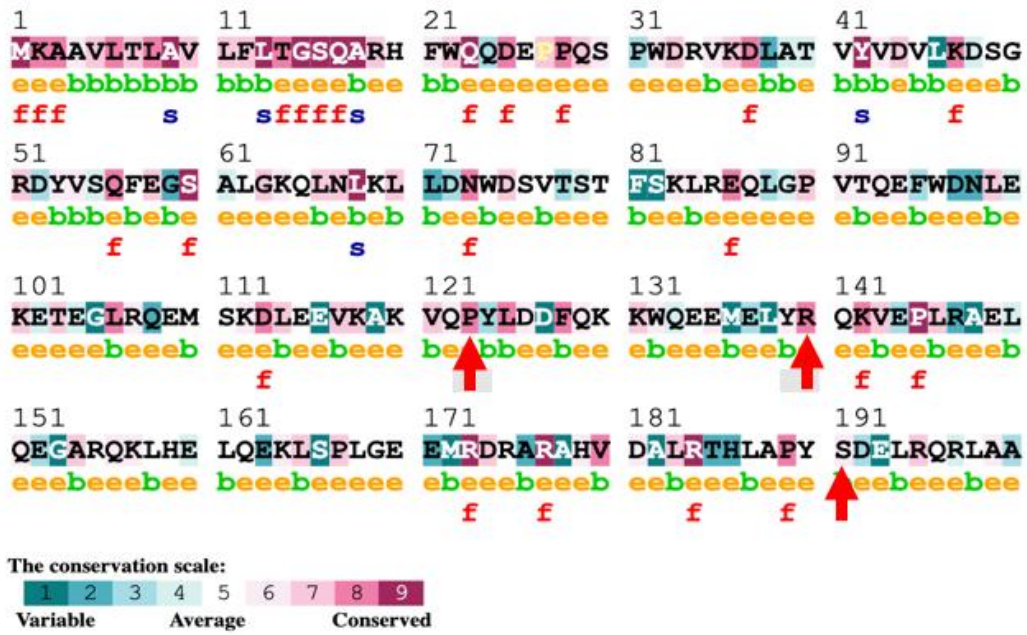

Figure 1. Consurf profile of apolipoprotein A1 (APOA1). Value 1 indicates a high variability region. The value increases as the region becomes more conserved, up to value 9. The position values of four of the deleterious SNVs considered to be deleterious by at least three programs in each category (sequence homology-based methods, supervised learning methods and consensus-based methods) are indicated by red arrows. e – denotes an exposed residue according to the neural-network algorithm; b – indicates a buried residue according to the neural-network algorithm; f – indicates a predicted functional residue (highly conserved and exposed); s – indicates a predicted structural residue (highly conserved and buried).

IV. CONCLUSION

In this study, using missense SNV impact prediction programs and multiple alignment for evolutionary insights, we observed that four SNVs could have a harmful effect on the APOA1 protein. For this, 12 *in silico* SNV impact prediction programs were used to investigate the protein coded by the *APOA1* gene. Results showed that from the 15 missense SNVs initially studied, four (Pro123Ser, Arg140Leu, Ser191Asn and Ser191Arg) were considered as deleterious by at least three programs in each category studied. Among these, two SNVs (Pro123Ser and Arg140Leu) are located in highly conserved regions of the *APOA1* gene.

REFERENCES

[1]. Breslow, J. L., Zannis, V. I., SanGiacomo, T. R., Third, J. L., Tracy, T., & Glueck, C. J. (1982). Studies of familial type III hyperlipoproteinemia using as a genetic marker the apoE phenotype E2/2. *Journal of Lipid Research*, *23*(8), 1224–1235. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7175379

[2]. Bromberg, Y., Yachdav, G., & Rost, B. (2008). SNAP predicts effect of mutations on protein function. *Bioinformatics (Oxford, England)*, *24*(20), 2397–2398. https://doi.org/10.1093/bioinformatics/btn435

[3]. Capriotti, E, Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics (Oxford, England)*, *22*(22), 2729–2734. https://doi.org/10.1093/bioinformatics/btl423

[4]. Capriotti, Emidio, 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*(Suppl 3), S6. https://doi.org/10.1186/1471-2164-14-S3-S6

[5]. Capriotti, Emidio, Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, *33*(SUPPL. 2). https://doi.org/10.1093/nar/gki375

[6]. Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., … Ben-Tal, N. (2013). ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Israel Journal of Chemistry*, *53*(3–4), 199–206. https://doi.org/10.1002/ijch.201200096

[7]. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PloS One*, *7*(10), e46688. https://doi.org/10.1371/journal.pone.0046688

[8]. Dastani, Z., Dangoisse, C., Boucher, B., Desbiens, K., Krimbou, L., Dufour, R., … Marcil, M. (2006). A novel nonsense apolipoprotein A-I mutation (apoA-IE136X) causes low HDL cholesterol in French Canadians. *Atherosclerosis*, *185*(1), 127–136. https://doi.org/10.1016/j.atherosclerosis.2005.05.028

[9]. De Baets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J., … Rousseau, F. (2012). SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Research*, *40*(Database issue), D935-9. https://doi.org/10.1093/nar/gkr996

[10]. Dorn, M., E Silva, M. B., Buriol, L. S., & Lamb, L. C. (2014). Three-dimensional protein structure prediction: Methods and

computational strategies. *Computational Biology and Chemistry*, Vol. 53, pp. 251–276. https://doi.org/10.1016/j.compbiolchem.2014.10.001

[11]. Ertek, S. (2017). High-density Lipoprotein (HDL) Dysfunction and the Future of HDL. *Current Vascular Pharmacology*, *16*(5), 490–498. https://doi.org/10.2174/1570161115666171116164612

[12]. Fokkema, I. F. A. C., Taschner, P. E. M., Schaafsma, G. C. P., Celli, J., Laros, J. F. J., & den Dunnen, J. T. (2011). LOVD v.2.0: The next generation in gene variant databases. *Human Mutation*, *32*(5), 557–563. https://doi.org/10.1002/humu.21438

[13]. Franceschini, G., Sirtori, M., Gianfranceschi, G., & Sirtori, C. R. (1981). Relation between the HDL apoproteins and AI isoproteins in subjects with the AIMilano abnormality. *Metabolism*, *30*(5), 502–509. https://doi.org/10.1016/0026-0495(81)90188-8

[14]. González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American Journal of Human Genetics*, *88*(4), 440–449. https://doi.org/10.1016/j.ajhg.2011.03.004

[15]. Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073–1081. https://doi.org/10.1038/nprot.2009.86

[16]. Ladias, J. A. A., Kwiterovich, P. O., Smith, H. H., Karathanasis, S. K., & Antonarakis, S. E. (1990). Apolipoprotein A1 Baltimore (Arg10→Leu), a new ApoA1 variant. *Human Genetics*, *84*(5), 439–445. https://doi.org/10.1007/BF00195816

[17]. Liu, M., Wang, L., Sun, X., & Zhao, X. (2014). Investigating the impact of Asp181 point mutations on interactions between PTP1B and phosphotyrosine substrate. *Scientific Reports*, *4*, 5095. https://doi.org/10.1038/srep05095

[18]. Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., … Thomas, P. D. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Research*, *33*(Database issue), D284-8. https://doi.org/10.1093/nar/gki078

[19]. Miller, M. P., & Kumar, S. (2001). Understanding human disease mutations through the use of interspecific genetic variation. *Human Molecular Genetics*, *10*(21), 2319–2328. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11689479

[20]. Ng, P. C., & Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein function. *Annual Review of Genomics and Human Genetics*, *7*, 61–80. https://doi.org/10.1146/annurev.genom.7.080505.115630

[21]. Niroula, A., Urolagin, S., & Vihinen, M. (2015). PON-P2: Prediction Method for Fast and Reliable Identification of Harmful Variants. *PloS One*, *10*(2), e0117380. https://doi.org/10.1371/journal.pone.0117380

[22]. Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*, *39*(17), e118. https://doi.org/10.1093/nar/gkr407

[23]. Shen, J., Deininger, P. L., & Zhao, H. (2006). Applications of computational algorithm tools to identify functional SNPs in cytokine genes. *Cytokine*, *35*(1–2), 62–66. https://doi.org/10.1016/j.cyto.2006.07.008

[24]. Strobl, W., Jabs, H. U., Hayde, M., Holzinger, T., Assmann, G., & Widhalm, K. (1988). Apolipoprotein A-I (Glu 198→lys): A mutant of the major apolipoprotein of high-density lipoproteins occurring in a family with dyslipoproteinemia. *Pediatric Research*, *24*(2), 222–228. https://doi.org/10.1203/00006450-198808000-00017

[25]. Utermann, G., Pruin, N., & Steinmetz, A. (1979). Polymorphism of apolipoprotein E. III. Effect of a single polymorphic gene locus on plasma lipid levels in man. *Clinical Genetics*, *15*(1), 63–72.

[26]. van der Vorst, E. P. C. (2020). High-Density Lipoproteins and Apolipoprotein A1. *Sub-Cellular Biochemistry*, Vol. 94, pp. 399–420. https://doi.org/10.1007/978-3-030-41769-7_16

[27]. Yates, C. M., Filippis, I., Kelley, L. A., & Sternberg, M. J. E. (2014). SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *Journal of Molecular Biology*, *426*(14), 2692–2701. https://doi.org/10.1016/j.jmb.2014.04.026

[28]. Yates, C. M., & Sternberg, M. J. E. (2013). The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. *Journal of Molecular Biology*, *425*(21), 3949–3963. https://doi.org/10.1016/j.jmb.2013.07.012

S. A. Alencar, et. al. "A preliminary Analysis of the Impact of Missense Variants of Unknown Significance in the APOA1 Gene using prediction softwares." *The International Journal of Engineering and Science (IJES)*, 9(8), (2020): pp. 47-51.