In silico Analysis of Non Synonymous Single Nucleotide Polymorphisms of Human *LPA g*ene associated with Ischemic Heart Disease

Kulkarni Vineet[1], Shetty Sukanya[2], Bhandary Roopa[2*] and Mary Yanglem[2]
1. Tejasvini Institute of Allied Health Sciences, Mangaluru, Karnataka, INDIA
2. Department of Biochemistry, KS Hegde Medical Academy, Nitte (Deemed to be University), Mangaluru, Karnataka, INDIA
*bhandarybio@nitte.edu.in

Abstract
Ischemic heart disease (IHD) is a serious medical condition that has been linked to higher levels of Lipoprotein (a) in the blood. However, the genetic factors that lead to this increase in Lipoprotein (a) and the subsequent risk of IHD, are not well understood. To shed more light on this issue, this study aims to use computational methods to investigate the role of functional single nucleotide polymorphism(SNP) in Lp(a)(LPA) gene expression.Functional SNPs for the LPA gene were retrieved from the db SNP database. Two in silico tools, SIFT and SNAP 2, were used to identify the most damaging SNPs in the LPA gene. The Mupro web tool was used to predict protein stability changes resulting from single-site mutations in the APOA protein. The web tool ConSurf was used to determine the evolutionary conservation analysis of functional SNPs in LPA. The project HOPE software was used to predict protein structure and to analyse the effects of the mutations.29 missense SNPs were selected for this study using the SIFT database.

According to the MUPro Server, 27 SNPs were found to decrease protein stability (G0.5). Among the 29 SNPs investigated, 6 were found to be average, 11 were conserved and 8 were not conserved evolutionarily. According to HOPE analysis, all SNPs were found to be damaging which could alter the properties of amino acids in different Kringle regions. This study provides evidence that polymorphism in the LPA gene can impact its cellular function and serum Lp(a) level.

Keywords: Ischemic heart disease, lipoprotein(a), single nucleotide polymorphism

Introduction

Cardiovascular diseases (CVD) are a significant global health concern, responsible for an alarming 17.9 million deaths each year. Ischemic heart disease (IHD) is one such condition that develops gradually due to the growth of lesions in the coronary arteries, leading to insufficient oxygen supply to the heart[4]. Known risk factors including physical inactivity, smoking, alcohol consumption, obesity and diabetes mellitus, failed to assess IHD risk completely[3]. Lipoprotein(a) [Lp(a)], a glycoprotein that forms a particle similar to low-density lipoprotein(LDL), is believed to play a crucial role in preventing excessive bleeding during childbirth[14]. Apolipoprotein(a) gene (*LPA*) regulates the serum level of Lp(a) which is highly heritable. The *LPA* gene has unique Kringle domains (KI to KV) and a serine protease domain and the size of the apo(a) isoform is inversely correlated with the serum Lp(a) concentration[9]. Larger apo(a) isoforms are more susceptible to degradation, resulting in lower serum Lp(a) levels[7,10,12].

Recent studies have highlighted the vital role of single nucleotide polymorphisms (SNPs) in the *LPA* gene in regulating serum Lp(a) levels and increasing the risk of cardiovascular disease across ethnicities[1,11].

Computational studies using *in silico* analysis can help to identify potentially disease-causing or IHD risk-associated *LPA* SNPs by assessing their impact on gene structure, protein function, regulatory elements and interactions within biological pathways. Therefore, the primary objective of this study is to identify and to assess the role of functional *LPA* SNPs in gene expression, which could lead to the development of new treatments and strategies to manage IHD and to improve cardiovascular health outcomes.

Material and Methods

Selection of Functional SNPs in Human *LPA* Gene: To investigate the functional SNPs specific to the LPA gene, we utilized the dbSNP database (https://www.ncbi.nlm.nih.gov/snp/). After filtering through a total of 5,600 SNPs, we shortlisted 29 based on their allele frequency. It should be emphasized that all 29 SNPs were located in the exonic region of the *LPA* gene.

Prediction of Damaging SNPs: Current study employed two *in silico* tools, SIFT (https://sift.bii.a-star.edu.sg/) and SNAP 2 (https://github.com/KorfLab/SNAP) to predict the most damaging SNPs in the *LPA* gene. We inputted substituting amino acids along with their positions and FASTA sequences of Apo(a) proteins into each web tool.

Effect of SNPs on Structure and Function of Apo(a) Protein: To determine the deleterious effect of SNPs on Apo(a) protein expression, this study used Mutpred 1 (http://mutpred.mutdb.org/), a web-based tool that takes into account the structural and functional properties of amino acids.

Impact of SNPs on Protein Stability: This study used the Mupro web tool (https://mupro.proteomics.ics.uci.edu/) to

predict protein stability changes for single-site mutations in Apo(a) protein.

Conservation Analysis of Functional SNPs: Current study performed an evolutionary conservation analysis of functional SNPs of *LPA* gene using the web tool Con Surf (http://consurf.tau.ac.il/). The web tool measures the conservation score of amino acid position based on the phylogenetic relation between homologous sequences of apo(a) protein.

Post-Translational Modification Prediction: To identify potential sites in apo(a) sequence , present study determined phosphorylation and methylation sites by using the GPS MSP server (https://msp.biocuckoo.org/).

Prediction of Protein Structure and Mutant Analysis: This study valuated the functional impact of SNPs on point mutation and structural integrity of apo(a) protein using the HOPE (Have (y) Our Protein Explained) web server (https://www3.cmbi.umcn.nl/hope/method/). The HOPE server predicts protein structure and mutation analysis by acquiring data from sources such as the UniProt database.

Results

Retrieval of Functional SNPs from the NCBI SNP Database: The NCBI SNP database reported a total of 53,020 SNPs in the human *LPA* gene. Among these SNPs, 9 were benign, 1 was likely benign, 1714 were missense, 49699 were intronic and 477 were non-coding transcript. Additionally, 1,120 were synonymous SNPs. Since damaging SNPs can affect the structural and functional aspects of proteins, 29 missense SNPs were selected for the present study.

Prediction of Damaging SNPs: To identify potentially deleterious SNPs, 1714 missense SNPs were submitted to the Sorting Intolerant from Tolerant (SIFT) web-based tool. SIFT predicted that 29 SNPs can impact apo(a) protein expression (≤0.05) while the remaining SNPs were either tolerant or benign.

To further assess the impact of SNPs, the SNPs were analyzed using SNAP 2. The results from the SNAP 2 web tool revealed that among 29 LPA SNPs, 18 were found to be effective and 11 were neutral (Table 1).

Table 1
Assessment of functional impact of SNPs by SIFT and SNAP2 tools

SNP ID	Location	Amino acid substitution	SNAP2	SIFT
rs200561706	160594009	Ser1193Phe	Effect	Deleterious
rs200802664	160606480	Gln928Glu	Neutral	Deleterious
rs201200716	160586532	Thr1349Met	Effect	Tolerated
rs201244072	160635204	Gln332Glu	Effect	Tolerated
rs201290459	160577280	Ser1496ile , ser1496asn	Effect , effect	Deleterious
rs369875346	160589571	Thr1310Ile	neutral	Deleterious
rs372640159	160611629	Thr846Ser	Neutral	Deleterious
rs374480077	160548522	Pro1704Leu	Effect	Deleterious
rs374776600	160547908	Arg1729Trp	effect	Tolerated
rs535276500	160594096	Gly1164Ala	neutral	Tolerated
rs544137228	160594115	Pro1158Ala,Pro1158Thr	Effect,effect	Deleterious
rs544706080	160545506	Asn1778Asp	effect	Deleterious
rs554026616	160577182	Met1529Leu	Effect	Tolerated
rs556277165	160547802	Gly1764Asp	effect	Tolerated
rs557039318	160599537	His1084Tyr,His1084Asn	Effect,effect	Deleterious
rs558138162	160606635	Arg876Lys	effect	Deleterious
rs559005335	160664177	Phe13Tyr	effect	Deleterious
rs559188603	160586517	Met1354Thr	effect	Deleterious
rs566459406	160606633	Asn877Tyr	effect	Deleterious
rs572190082	160640694	Pro236Ser	effect	Deleterious
rs572818292	160557454	Glu1583Asp	effect	Deleterious
rs765765045	160635174	Ala342Thr	neutral	Deleterious
rs770691046	160635260	Ala313Gly,Ala313Asp	neutral,neutral	Tolerated
rs772415624	160545526	Arg1771His	neutral	Tolerated
rs1355346409	160640699	Pro234Leu,Pro234Arg	effect,effect	Deleterious
rs1428658532	160633829	Thr387Ala	neutral	Deleterious
rs1484919544	160640667	Ala245Thr	neutral	Deleterious
rs1582891859	160633814	Gln392Tyr,Gln392Glu	neutral,neutral	Deleterious
rs1672519011	160640805	Ala199Ser,Ala199Pro	Neutral,effect	Tolerated

Table 2
Evolutionary conservation analysis of SNPs in LPA gene

SNP ID	Evolutionary conservation analysis
rs200561706	Insufficient data
rs200802664	Average (5)
rs201200716	Conserved (7)
rs201244072	Not conserved (4)
rs201290459	Insufficient data
rs369875346	Conserved (7)
rs372640159	Average (5)
rs374480077	Not conserved (3)
rs374776600	Conserved (8)
rs535276500	Not conseved (1)
rs544137228	Average (5)
rs544706080	Conserved
rs554026616	Conserved (8)
rs556277165	Average (5)
rs557039318	Conserved (9)
rs558138162	Conserved (7)
rs559005335	Not conserved
rs559188603	Conserved (9)
rs566459406	Conserved (7)
rs572190082	Not conserved (3)
rs572818292	Not conserved (1)
rs765765045	Not conserved (4)
rs770691046	Conserved (9)
rs772415624	Average (5)
rs1355346409	Conserved (8)
rs1428658532	Not conserved
rs1484919544	Conserved (9)
rs1582891859	Insufficient data
rs1672519011	Insufficient data

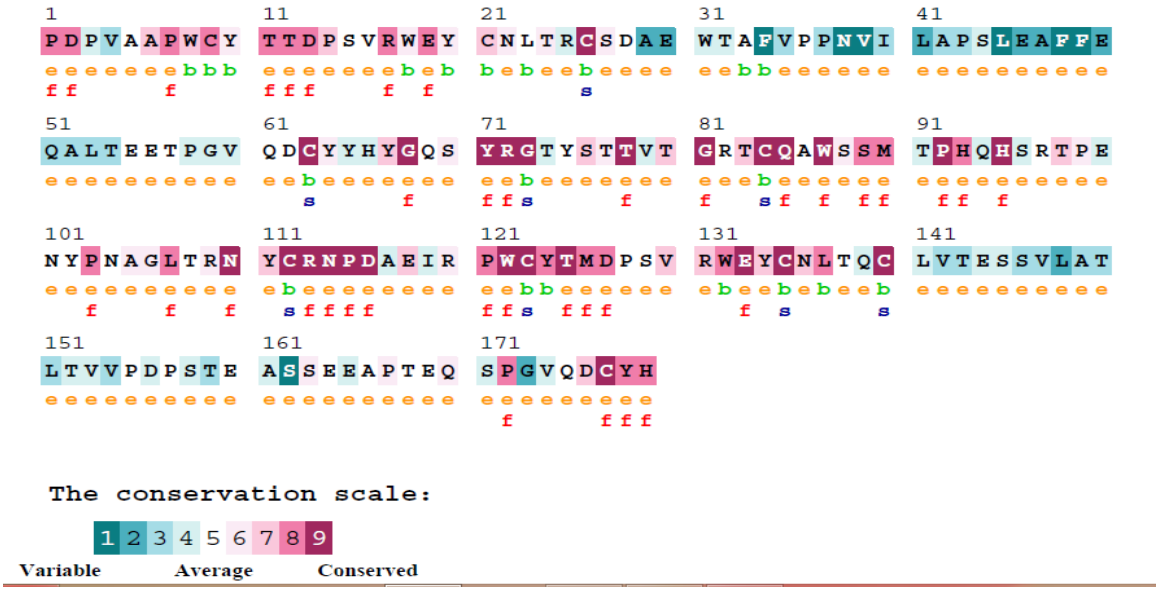

Fig. 1: Evolutionary conservation analysis of SNPs by con surf tool.According to this web tool, "f" is denoted as functional residue and "s" is denoted as structural residue. The continuous conservation scores are divided into a discrete scale of nine grades for visualization, from the most variable positions (grade 1) colored turquoise, through intermediately conserved positions (grade 5) colored white, to the most conserved positions (grade 9) colored maroon

Table 3
Effect of SNPs on apo (a) protein stability

SNP ID	MU PRO (Protein stability)
rs200561706	Decrease (G- -0.21)
rs200802664	Decrease (G - 0.35)
rs201200716	Decrease (G- 0.45)
rs201244072	Decrease (G- 0.41)
rs201290459	Decrease (G - 0.13)
rs369875346	Decrease (G -0.19)
rs372640159	Decrease (G- -0.36)
rs374480077	Decrease (G- 0.7)
rs374776600	Decrease (G- 0.4)
rs535276500	Decrease (G- 0.8)
rs544137228	Decrease (G - 0.91)
rs544706080	Decrease (G - -0.61)
rs554026616	Decrease (G - 0.81)
rs556277165	Increase (G -0.129)
rs557039318	Decrease (G- -1.5)
rs558138162	Decrease (G- -1.5)
rs559005335	Decrease (-0.8)
rs559188603	Decrease (-0.8)
rs566459406	Decrease (-.71)
rs572190082	Decrease (-0.4)
rs572818292	Decrease (-1.7)
rs765765045	Decrease (-1.4)
rs770691046	Increase (G- 0.18)
rs772415624	Decrease (G- 0.3)
rs1355346409	Decrease (-0.3)
rs1428658532	Decrease (-0.26)
rs1484919544	Decrease (-1.3)
rs1582891859	Decrease (G- -0.25)
rs1672519011	Decrease (G - 0.37)

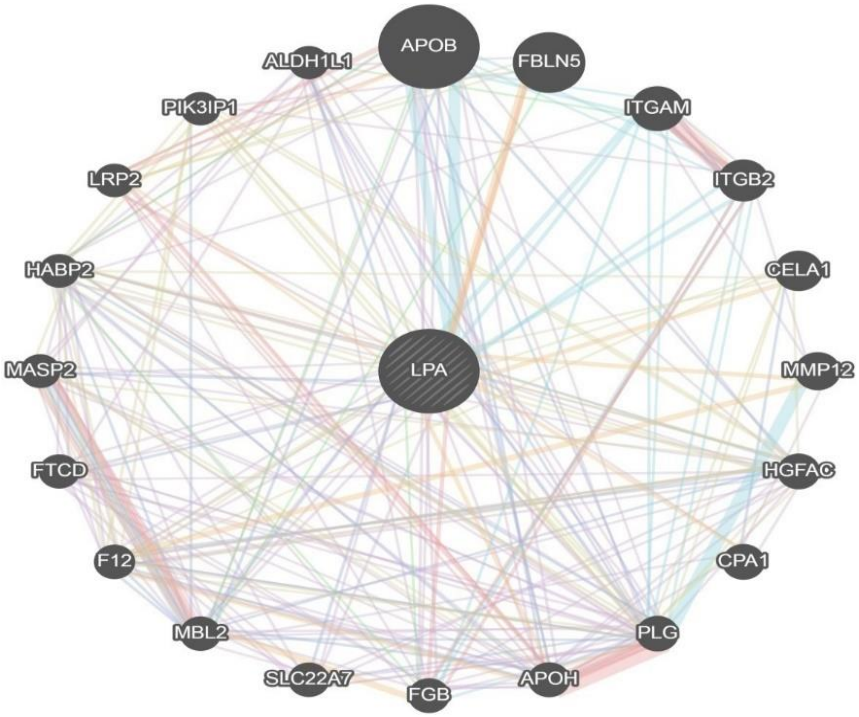

Fig. 2: Gene gene interactions analysed using GeneMania where each node in the network represents one gene. It is a biological network integration for gene prioritization and is useful in predicting gene function

Table 4
Post translation modification analysis by GPS MSP

SNP ID	Methylation	Phosphorylation
rs200561706	Methylated(23)	Phosphorylated
rs200802664	Methylated (15)	Phosphorylated
rs201200716	Non methylated(8.8)	Non Phosphorylated
rs201244072	Non methylated(7.65)	Non Phosphorylated
rs201290459	Methylated (37.96)	Phosphorylated
rs369875346	Methylated (24.80)	Phosphorylated
rs372640159	Non methylated (48.42)	Non Phosphorylated
rs374480077	Methylated (8.88)	Phosphorylated
rs374776600	Methylated (108.09)	Phosphorylated
rs535276500	Methylated (24.31)	Phosphorylated
rs544137228	No effect	Phosphorylated
rs544706080	Non methylated (27.27)	Non Phosphorylated
rs554026616	Methylated(62.71)	Non Phosphorylated
rs556277165	Non methylated (21.99)	Non Phosphorylated
rs557039318	Methylated(39.97)	Non Phosphorylated
rs558138162	Non methylated (10.90)	Phosphorylated
rs559005335	Non methylated (3.75)	Phosphorylated
rs559188603	Non methylated (7.65)	Non Phosphorylated
rs566459406	Non methylated (10.90)	Phosphorylated
rs572190082	Methylated (15.86)	Phosphorylated
rs572818292	Non methylated(7.35)	Phosphorylated
rs765765045	Non methylated (7.65)	Phosphorylated
rs770691046	Non methylated (9.11)	Non Phosphorylated
rs772415624	Non methylated (27.27)	Non Phosphorylated
rs1355346409	methylated (15.86)	Phosphorylated
rs1428658532	Non methylated (48.42)	Phosphorylated
rs1484919544	methylated (15.86)	Phosphorylated
rs1582891859	Non methylated (24.44)	Non Phosphorylated
rs1672519011	Non methylated (9.11)	Non Phosphorylated

Evolutionary Conservation Analysis: Evolutionary conservation analysis in the *LPA* gene was performed using Consurf. On a scale from 1-10, the Consurf server predicts whether the given SNP is in a conserved (8-10) or non-conserved position (1-5). Among the 29 SNPs investigated, 6 SNPs were average, 11 SNPs were conserved, 8 SNPs were non-conserved and the remaining SNPs could not be analyzed due to insufficient data (Table 2) (Fig. 1).

Impact of SNPs on Protein Stabilization: To evaluate the effects of SNPs on protein stability, the bioinformatics tool MU pro was utilized. According to the findings from the MU pro server, 27 SNPs were reported to have decreased protein stability ($\Delta G \leq 0.5$) while 2 SNPs showed an increase in protein stability (Table 3).

Prediction of Post-Translational Modification by GPS MSP: The study analyzed methylation and phosphorylation using the GPS MSP web tool. The web tool predicted methylation sites in 12 out of 29 amino acid residues and phosphorylation in 18 amino acid residues. Twelve amino acid residues were reported to induce both phosphorylation and methylation (Table 4).

Impact of Amino Acid Substitution on APO(a) Protein from HOPE: Amino acid analysis was performed using the HOPE web tool. According to HOPE analysis, all SNPs were found to be damaging, potentially altering the properties of amino acids in different Kringle regions (Table 5).

Gene-Gene Interaction : GeneMANIA (https://genemania.org/) was employed to analyze the interaction of the *LPA* gene with other genes. The network analysis revealed that the LPA gene has physical interactions with APOB, FBLN5, PLG and APOH (Fig. 2).

Discussion

IHD is a significant contributor to mortality among cardiovascular diseases. SNPs play a crucial role in determining disease prognosis and comprehending the biological pathways that cause disease progression in individuals of diverse racial and ethnic backgrounds[6].

Table 5
Project HOPE analysis of *LPA* SNPs

SNP ID	Project HOPE analysis
rs200561706	Kringle-33(damaging)
rs200802664	Kringle-30(damaging)
rs201200716	Kringle-32(damaging)
rs201244072	Kringle-1(damaging)
rs201290459	Kringle-36(damaging)
rs369875346	Kringle-3(damaging)
rs372640159	Kringle-1(damaging)
rs374480077	Kringle-37(damaging)
rs374776600	Kringle38(damaging)
rs535276500	Kringle-33(damaging)
rs544137228	Kringle-33(damaging)
rs544706080	Kringle-38(damaging)
rs554026616	Kringle-36(damaging)
rs556277165	Kringle-38(damaging)
rs557039318	Kringle-32(damaging)
rs558138162	Kringle-30(damaging)
rs559005335	Kringle-31(damaging)
rs559188603	Kringle-34(damaging)
rs566459406	Kringle-30(damaging)
rs572190082	Kringle-1(damaging)
rs572818292	Kringle-36(damaging)
rs765765045	Kringle-1(damaging)
rs770691046	Kringle-1(damaging)
rs772415624	Kringle-38(damaging)
rs1355346409	Kringle-1(damaging)
rs1428658532	Kringle-2(damaging)
rs1484919544	Kringle-2(damaging)
rs1582891859	Kringle-2(damaging)
rs1672519011	Kringle-1(damaging)

This knowledge can aid in providing patients with personalized and effective treatment options. The present study aimed to investigate the effect of *LPA* SNPs on apo(a) expression which may impact serum Lp(a) concentration and IHD risk using various bioinformatics methods. SNPs were selected based on the NCBI database, allele frequency (0.4–1) and location (exonic region) in the *LPA* gene.

To enhance prediction accuracy, the study utilized the SNAP2 web server which assessed the functional effect of mutations in the amino acid sequence, distinguishing between effective and neutral variants. The findings from the SNAP2 web tool indicated that out of 29 *LPA* SNPs, 18 were effective and 11 were neutral respectively. Non-synonymous mutations may cause a shift in protein stability which can alter its functionality. According to the findings from the MU pro server, 27 SNPs were reported to have decreased protein stability ($\Delta G \leq 0.5$) whereas 2 SNPs showed an increase in protein stability.

For proteins to function normally, amino acid residues must have undergone evolutionary conservation[9]. However, it was hypothesized that these conserved sites are more likely to experience harmful mutations[5]. The 29 SNPs examined by Consurf revealed that *LPA* contained 11 SNPs in the conserved region. Under typical physiological circumstances, post-translational modifications (PTMs) may significantly contribute to the structural changes of proteins in various metabolic activities. Genetic research has shown evidence that functional SNPs may modify proteins post-translationally, changing their structural integrity. Proteins can have methyl and phosphoryl groups added to them as a result of post-translational modifications which can regulate their function[13].

In the last 20 years, proteomic research has confirmed that methylation and phosphorylation play a crucial role in many biological processes[8]. Present study has analyzed SNPs in a single gene with a limited number of bioinformatics tools which was one of the limitations. Large *in silico* studies with multiple gene analysis might provide more insights in understanding the molecular mechanisms in the progression of IHD complications.

The study found that among the SNPs examined, some were effective and some were neutral and non-synonymous

mutations might cause a shift in protein stability which can alter its functionality. The study also found that post-translational modifications may significantly contribute to the structural changes of proteins in various metabolic activities. These findings suggest that personalised treatment options based on SNPs can be beneficial in managing IHD complications. However, further research is needed to understand the molecular mechanisms involved in the progression of IHD complications.

Conclusion

Numerous bioinformatics tools were employed to identify deleterious SNPs in the *LPA* gene. A total of 29 non-synonymous SNPs were identified using diverse computational approaches. The presence of polymorphisms in the *LPA* gene may significantly impact its cellular function and serum Lp(a) levels. The identification of deleterious SNPs in the *LPA* region can be advantageous in predicting therapeutic strategies. *In silico* approaches may be utilized for individuals with damaging mutations in the *LPA* gene, thereby mitigating the risk of developing IHD.

References

1. Anu R., Kaur C., Bhavsar P. and Prasad J., Genotyping of Lipoprotein(a) Gene Variants in Coronary Artery Disease in Indian Population, *IJMB*, **25**, 76-82 **(2021)**

2. Beaney K.E. et al, SNP coronary heart disease gene score profile in subjects with type 2 diabetes: the coronary heart disease risk in type 2 diabetes (CoRDia study) study baseline characteristics, *Cardiovasc Diabetol*., **15**, 1-8 **(2016)**

3. Braffett B.H., Bebu I., El G.L., Cowie C.C., Sivitz W.I., Pop B.R., Larkin M.E., Gubitosi K.A. and Nathan D.M., Cardiometabolic Risk Factors and Incident Cardiovascular Disease Events in Women vs Men With Type 1 Diabetes, *JAMA Netw Open*, **5(9)**, 1-7 (**2022)**

4. Cegla J. et al, HEART UK consensus statement on Lipoprotein(a): A call to action, *Atherosclerosis*, **291**, 62–70 **(2019)**

5. Celniker G. et al, ConSurf: using evolutionary data to raise testable hypotheses about protein function, *Isr. J. Chem.*, **53**, 199–206 **(2013)**

6. Filipovic N. et al, *In Silico* Clinical Trials for Cardiovascular Disease, *J. Vis. Exp.*, **183**, 51-58 **(2022)**

7. Kamstrup P., Lipoprotein(a) and Cardiovascular Disease, *Clin. Chem*, **67**, 154–166 **(2021)**

8. Kim Y., Kang C., Min B. and Yi G.S., Detection and analysis of disease associated single nucleotide polymorphism influencing post-translational modification, *BMC Medical Genom*, **8**, 2-8 **(2015)**

9. Kronenberg F., Lipoprotein(a) in atherosclerotic cardiovascular disease and aortic stenosis: a European Atherosclerosis Society consensus statement, *Eur. Heart J*, **43**, 3925–3946 **(2022)**

10. Kronenberg F., Human Genetics and the Causal Role of Lipoprotein(a) for Various Diseases, *Cardiovasc Drugs Ther*, **30**, 87–100 **(2016)**

11. Lanktree M.B., Anand S.S., Yusuf S. and Hegele R.A., Comprehensive Analysis of Genomic Variation in the *LPA* Locus and Its Relationship to Plasma Lipoprotein(a) in South Asians, Chinese and European Caucasians, *Circ. Cardiovasc, Genet*, **3**, 39–46 **(2010)**

12. Parish S. et al, Impact of Apolipoprotein(a) Isoform Size on Lipoprotein(a) Lowering in the HPS2-THRIVE Study, *Circ. Cardiovasc. Genet*, **11**, 65-69 **(2018)**

13. Venselaar H., Te Beek T.A., Kuipers R.K.P., Hekkelman M.L. and Vriend G., Protein structure analysis of mutations causing inheritable diseases an e-Science approach with life scientist friendly interfaces, *BMC Bioinf*, **11**, 1–10 **(2010)**

14. Zubair H., Junaid I., Hongcai L. and Hou Z., Exploring the role of lipoprotein(a) in cardiovascular diseases and diabetes in Chinese population, *Int. J. Biol. Macromol*, **233**, 1-9 **(2023)**.

(Received 16th October 2024, accepted 15th November 2024)