CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Determination and Impact Analysis of Deleterious SNPs of ApoE

by

Shazina Pervaiz

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2021

I dedicate this thesis to my parents and my teachers.

CERTIFICATE OF APPROVAL

Determination and Impact Analysis of Deleterious SNPs of ApoE

by

Shazina Pervaiz

(MBS191002)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Syed Babar Jamal Bacha	NUMS, Rawalpindi
(b)	Internal Examiner	Dr. Sahar Fazal	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
April, 2021

Dr. Sahar Fazal
Head
Dept. of Biosciences & Bioinformatics
April, 2021

Dr. Muhammad Abdul Qadir
Dean
Faculty of Health & Life Sciences
April, 2021

Author's Declaration

I, **Shazina Pervaiz** hereby state that my MS thesis titled "**Determination and impact analysis of deleterious SNPs of ApoE**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Shazina Pervaiz)

Registration No: MBS191002

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Determination and impact analysis of deleterious SNPs of ApoE**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been dully acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Shazina Pervaiz)

Registration No: MBS191002

Acknowledgement

To begin with the best name of **Almighty Allah**, most gracious and merciful, whose bounties are unbounded, whose benevolence is everlasting, whose advantages are uncountable, whose being is eternal, whose mercy is unlimited, whose provisions are un-finishing and whose love is our existence, whose worship is our faith. Words appear too little to express my gratitude to Allah Almighty who has bestowed me with more than I deserve.

I would like to specifically express my sincere gratitude to Capital University of Science and Technology (CUST) Islamabad for providing me an opportunity to do MS Biosciences and achieve my aim to do better research. I would like to extend special thankfulness to my supervisor, **Dr. Syeda Marriam Bakhtiar**, for her regular aid, encouragement, and guidance throughout this thesis. Special gratitude to my friends and colleagues for assisting me and encouraging me.

In the end, I specific my gratitude to my dad and mom, and siblings for imparting me unfailing support and non-stop encouragement during my years of studies and writing thesis.

(Shazina Pervaiz)

Abstract

ApoE is an apolipoprotein that relates to lipid molecules it is responsible for lipoprotein mediated lipid delivery between organs and occurs through the plasma and interstitial fluids. It is an important member of plasma lipoproteins and performs synthesis, exchange, and removal of lipoproteins. The apolipoproteins undergo interactions with chylomicrons, very low-density lipoproteins (VLDL) and intermediate density lipoproteins (IDL) but have special binding affinity for high-density lipoproteins. The ApoE gene is positioned on chromosome 19q13.32, it is a polymorphic gene and has three different alleles. These alleles have variations in amino acids at position 112 and 158 due to this mutation, individuals with different genotypes have distinct plasma lipid levels. Many diseases like coronary artery disorder, Alzheimer's disease, and osteoporosis involve disturbances in lipid levels. So, ApoE is taken into consideration as a risk factor for these types of illnesses. This study aims to determine the deleterious SNPs of APoE among many reported SNP's of ApoE. The deleterious effects are examined through using specific bioinformatics tools along with Sift Server, Polyphen 2 etc. After the prediction and validation of secondary structure of mutated proteins, it is aligned with wild-type protein structure to investigate the variations. Interactors of ApoE are identified to determine all of the possible pathways related to the ApoE genotype. Then ApoE pathways are built to analyze the role of ApoE in specific metabolic, cardiovascular, and neurological disorders.

Contents

List of Figures

List of Tables

Abbreviations

HDL	High Density Lipoprotein
LDL	Low Density Lipoproteins
LDs	Lipid Droplets
VLDL	Very Low Density Lipoproteins

Chapter 1

Introduction

1.1 Background

Apoliprotein E (ApoE) is an important glycoprotein that has 34 kDa molecular weight [1]. It consists of 299 amino acids after the cleavage of 18 amino acids signal peptide [2]. It performs its role in lipid metabolism and delivery as a plasma lipoprotein and occurs in various organs and inside different tissues. It was discovered in the early 1970s It additionally acts as a ligand to bind with low density lipoprotein receptors (LDL) and ApoE receptors. In this way, it supports lipids clearance with the aid of binding lipids to specific receptors.

There are three isoforms of ApoE protein the most common isoform is ApoE3 and it is also taken into consideration as a wild protein while ApoE2 and ApoE4 are considered mutated proteins with a difference of amino acid at position 112 or 158 [3]. The ApoE gene is present on chromosome 19q13.32. It has three common alleles, e2, e3, and e4 due to which it is called polymorphic gene.The proteins encoded by them are Apo E2, E3, and E4, respectively [4].

The unicellular and multicellular organisms perform metabolism with the help of enzymes to convert complex molecules to simpler forms. This conversion of substances is essential for the survival of residing organisms. Distinctive pathways exist in organisms for the distribution of molecules at the cell and organismic

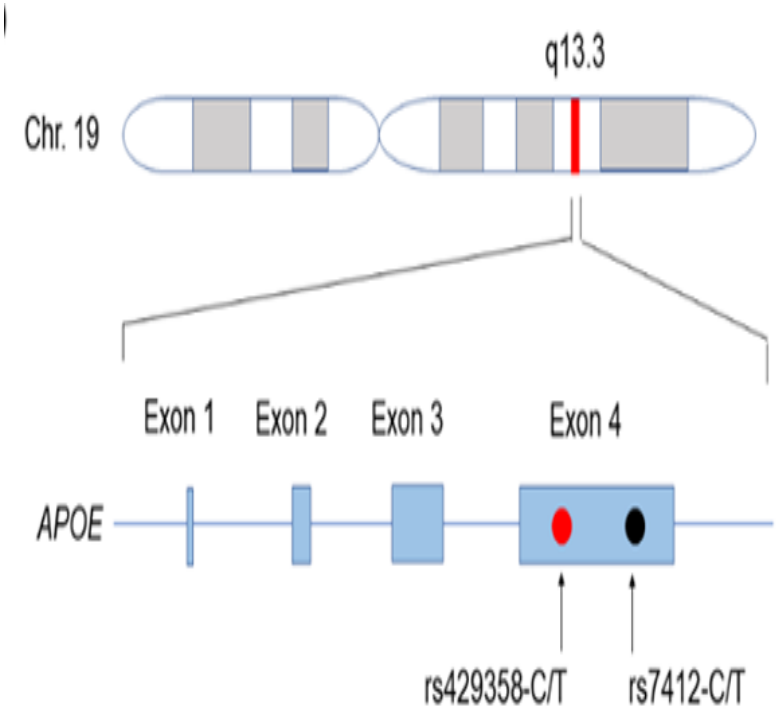

FIGURE 1.1: Structure of ApoE gene [5].

stage. To recognize the metabolic basis of these molecules, it is vital to understand their pathways. These pathways can be investigated to find a remedy for the different disorders. Some products may be removed or altered to reduce their detrimental effects. Moreover, the production of beneficial products can be enhanced after studying the pathways. The antibiotics can be investigated concerning specific pathways by developing new pathways after investigating existing pathways. The research on metabolic networks started in the opening quarter of the 1970s [6]. Pathways of cellular metabolism have the characteristic of plasticity. It isn't always so easy to decide the pathway via which a reactant is converted into products through the catalytic action of enzymes. Due to the complexity in pathways, practical genomics gives detailed information, for instance, DNA arrays provide analysis of the comprehensive patterns of gene expression. From these gene expressions, feasible phenotypes can be predicted. At present, three common methods for reading metabolic networks exist, all techniques are even though different they have the same motive of determining the metabolic pathways. The advantage of these strategies is that they don't require kinetic data. Although these strategies

are distinct from each other they are also associated with each other and are very beneficial for mapping biochemical networks without preconceptions. In the first technique, pathways are constructed by observing the successive reactants and products by linking them to complete a pathway. The second approach involves the complete admissible flux distributions that can be determined with the help of multiples of addition, subtraction data from flux distribution. With the additional advancement in research," elementary flux models" were proposed [7]. All living things need the energy to perform their everyday activities. Lipids are a rich source of reduced hydrocarbons. Organisms additionally have the potential to keep lipids due to their insufficient availability. Lipids are stored in cells and tissues in inert form. To stockpile lipids, all cells first change them into neutral lipids, consisting of triacylglycerol (TG) and sterol esters (SE) then they save them in their reservoirs. Lipids are stored in specialized intracellular organelles called lipid droplets (LDs) that are additionally called adiposomes, lipidbodies, or oil bodies.In addition to storage for energy, they're additionally a lipid reservoir for membrane synthesis (e.g., sterols, fatty acids, and phospholipids) [8].

Under a microscope, LDs appear as round structures and their diameter ranges from 0.1-5 mm in nonadipocytes and can be up to 100mm in white adipocytes. There is no membrane boundary around LDs like other sub-cellular organelles but a phospholipids monolayer surrounds it.These phospholipids include phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol, lysoPC, and lysoPE, and a number of them have specific houses [9]. Environmental conditions are not the same everywhere in the world or even in specific places conditions may fluctuate. So usually, those situations may turn out to be so tough for the survival of the organism which may be threatened due to the conditions of famine, drought, or by variations in reproductive stages. All living organisms adapt themselves for survival and specific behaviors exist for lipid metabolism. The lipids provide a rich source of strength, their storage can be a useful resource in harsh situations as an essential supply of energy [10]. The ApoE protein plays a considerable role in lipid metabolism and lipid delivery and alteration in a single nucleotide causes a mutation in amino acid which in the long run leads to different alleles of ApoE.

The structural change of ApoE proteins is responsible for the different behavior of these proteins in the direction of ApoE receptors and Low-Density Lipoprotein receptors. It has been found that people with ApoE3 homozygous genotypes have a normal level of total plasma and LDL cholesterol. ApoE2 organisms have high ApoE levels resulting in low plasma and LDL cholesterol level. ApoE4 individuals have low ApoE levels and overall high plasma and LDL cholesterol [11]. The studies with major attention on quantitative blood measures of lipid metabolism have resulted in association analysis. On the population level, $\sim$ 12%–20% of the interindividual variation exists inside the ApoE degree, in addition to 5%–8% of the interindividual variations arise in overall total cholesterol (TC) level, those variations are taken into consideration because of the six different genotypes produced through three alleles e2, e3, and e4 alleles. In a few regions, the difference in TG and/or HDL cholesterol (HDL-C) levels is likewise attributed to six ApoE genotypes [12]. The E2 homozygosity in type III hypolipoproteinemia and occurrence of the disorder in E 2/2 homozygotes help the study's outcomes.The gene E4 is taken into consideration as linked with hypercholesterolemia. Apo E4 is also related to kind V hyperlipidemia in a single observed organization however similarly studies have no longer established this affiliation [13].

1.2 Problem Statement

Analyzing the deleterious, detrimental effects of ApoE Single Nucleotide Polymorphisms in various cardiovascular, metabolic, and nervous disorders through pathway analysis.

1.3 Aims and Objectives

ApoE gene is involved in the pathogenesis of various diseases such as metabolic, cardiovascular, and neurological. Its role in these diseases indicates the significance of ApoE as a key regulator in multiple pathways. Protein interactions in

pathways are dependent on structural interactions between different proteins, ligands, metabolites, etc. Any mutation or polymorphism or genetic variation could result in a structural change of ApoE protein that affects its interactions. This study is designed to explore the genetic variations that could have an impact on ApoE protein structure and also elucidate that how these changes in structure will impact the overall efficiency and interaction of ApoE.

1.3.1 Major Objectives

1. To perform literature mining to determine the deleterious effects of ApoE gene SNPs in various diseases.

2. To predict and validate protein stability by constructing disease models through bioinformatics analysis using mutated ApoE protein sequences.

3. Constructing ApoE pathway to understand its role in different pathways.

4. Perform pathway analysis to determine the role of mutated ApoE protein in the pathophysiology of different diseases.

Chapter 2

Literature Review

ApoE has a critical role in metabolic and neurological pathways so any disturbance in ApoE can have an effect on these pathways and purpose a variety of neurological and metabolic problems [14]. ApoE is involved in lots of diseases concerning Cardiovascular diseases and Dementia. Diseases concerning ApoE gene and their relation with ApoE is given in the table 2.1.

Table 2.1: Disease associated with ApoE and their association with ApoE

Disease	**Association**	**Ref**
Ischemic Cerebrovascular Disease	Associated with higher ApoE4	[15]
Alzheimer's Disease	Associated with Higher ApoE4	[16]
Rheumatoid Arthritis	Associated with higher ApoE4	[17]
Metabolic syndrome in men	Associated with higher ApoE4	[18]
Polycystic Ovary	Higher ApoE3 in patients	[19]
Type2 Diabetes Mellitus	Higher ApoE3 in patients	[20]
Dementia	Associated with higher ApoE4	[21]
Atherosclerosis	Associated with higher ApoE4	[22]
Intraventricular Hemorrhage in infants	Associated with ApoE2 & ApoE4	[23]
Cardiovascular Disease	Associated with ApoE4	[24]

2.1 ApoE Protein

The primary ApoE protein shaped without delay because of translation includes 317 amino acids which later upon traslational modification becomes mature ApoE protein.Apolipoprotein E protein has largely been acknowledged for its function in plasma cholestrol and triglyceride metabolism through interacting with lipoprotein receptors [25]. Mature form of ApoE protein contains 299 amino acids. Among the three isoforms of protein,the difference is because of cysteine and arginine amino acid at position 112 and 158. ApoE 3 is the most commonly found form of ApoE and it contains cysteine and arginine position 112 and 158. ApoE2 has cysteine and ApoE 4 has arginine at both 112 and 158 positions [26]. Base differences in alleles of ApoE was shown in **Table 2.2**.

TABLE 2.2: Base differences in alleles of ApoE [5]

Allele	**E2**	**E3**	**E4**
Haplotype	rs429358-T rs7412-T	rs429358-T rs7412-C	rs429358-C rs7412-C
Residue	112-Cys	112-Cys	112-Arg
Combination	158-Cys	158-Arg	158-Arg

Consistent with the physical and chemical studies, it has been confirmed that ApoE protein has two domains that are separated by thrombin cleavage which is a fragment of a 22-kDa, residues 1-191, and a fragment of 10-kDa, residues 216-299, these two domains are folded independently [26][27].

Amino terminal domain is responsible for interaction between ApoE and low-density lipoprotein receptor [25][28]. And the carboxyl-terminal domain determines the binding with lipoprotein [28][29][30]. The X-ray Crystallography structure indicates that ApoE amino-terminal has an extended 4-helix bundle [26]. A unique feature related to ApoE is that one area can affect the functional abilities of another area. The carboxyl-terminal domain affects the binding ability of

two variations of ApoE with faulty binding capabilities. It has been proved by removing the carboxyl-terminal area in ApoE2 and detecting an increase within the binding activity of the defective version [31][32].

Like the carboxyl-terminal area, the amino-terminal domain also has the potential to affect the properties of the carboxyl-terminal area. ApoE3 indicates an excessive binding affinity for high-density lipoprotein and ApoE4 has an excessive affinity for binding with VLDL and IDL [28][33][34]. This distinction is taken into consideration to be due to area interaction that's a result of interchange between cysteine-arginine present at position 112 in the amino-terminal area.

ApoE3 indicates an excessive binding affinity for high-density lipoprotein and ApoE4 has an excessive affinity for binding with VLDL and IDL [28][33][34]. This distinction is taken into consideration to be due to area interaction that's a result of interchange between cysteine-arginine at position 112 in the amino-terminal area. This interchange is also considered to be a cause of a new isomer of ApoE protein and it determines the type of lipoproteins to which the carboxyl-terminal area binds [28].

In ApoE3 positive charge at position 112 is also taken into consideration as an element for the dedication of domain interaction. This consideration changed into tested by replacing the cysteine with a lysine residue. This change resulted in the alternation in lipoprotein nature from HDL to VLD/IDL.

Due to those observations, it became postulated that the exchange in lipoprotein possibilities and domain interaction was due to the exchange of cysteine-arginine amino acid at position 112 in carboxyl-terminal domain [28].

Moreover, it has also been studied that people with ApoE4 genotype have extra susceptibility for atherosclerosis compared to human beings with ApoE2 and ApoE3 because of an increase in plasma level of cholesterol and LDL [35].

It's important to be noted that this distinction of cholesterol and LDL is because of the alternate in domain interaction in ApoE4 and resulting change in nature of lipoproteins binding to ApoE. In recent years, ApoE4 has additionally been

taken into consideration as a threat thing for Alzheimer's disease [36]. Function of isoforms of protein is altered by change in its structure.

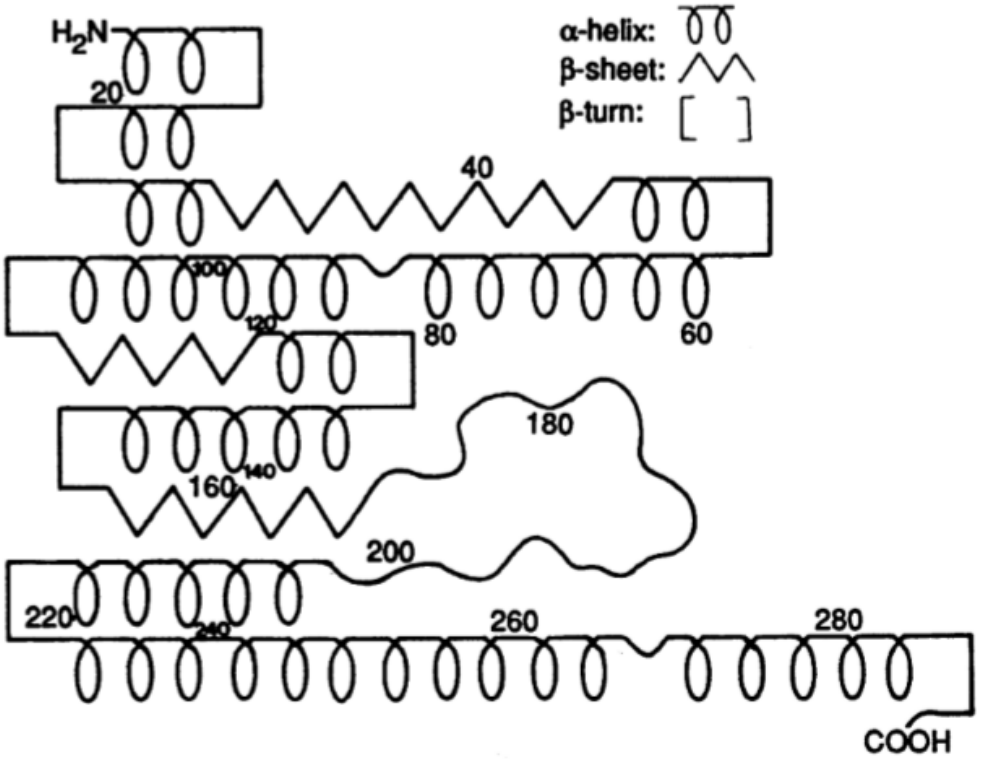

FIGURE 2.1: Secondary Structure of ApoE Protein [1].

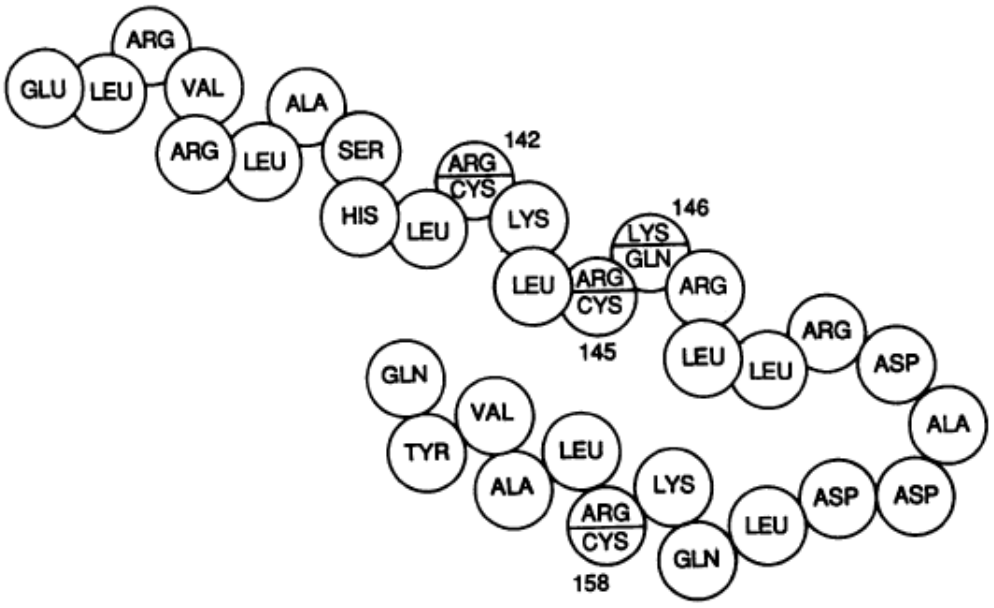

FIGURE 2.2: Aminoacids of ApoE in the nearby region of receptors binding domain

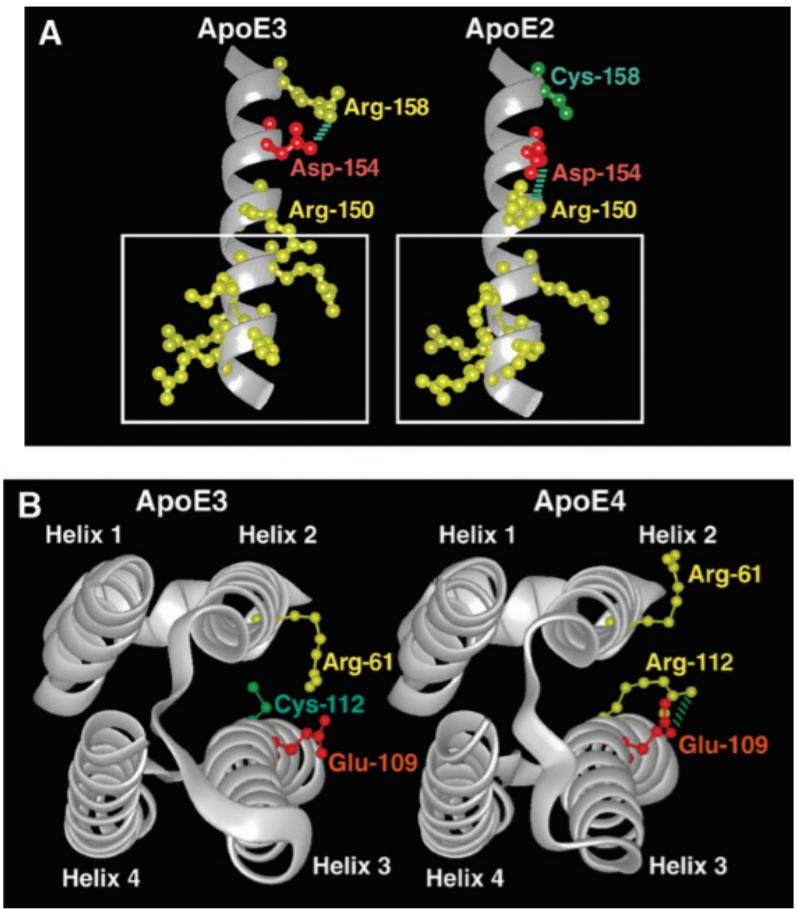

FIGURE 2.3: A). Secondary Structure of ApoE3 and ApoE2. B) Secondary structure of ApoE3 and ApoE4 [37].

2.2 ApoE Gene

ApoE gene is located on chromosome 19 (19q13.2) in the form of cluster at the side of the ApoC-I , ApoC-I', and ApoC-II genes [38]. The gene for LDL receptor is also located on this chromosome [1]. ApoE is polymorphic gene(gene having more than one alleles). Alleles of ApoE are $\varepsilon 2$, $\varepsilon 3$, and $\varepsilon 4$ which code for 3 isoforms of ApoE protein (E2, E3, and E4). Existence of any two alleles of the ApoE among its three naturally occurring alleles brings about six different genotypes with 3 homozygous (E2/E2, E3/E3, and E4/E4) and 3 heterozygous (E2/E3, E2/E4, and E3/E4)[39]. The length of ApoE gene is 3.7 kilobases. It has four exons [1]. The promotor sequence is TATAATT which is located on -30 position(30 basepairs up streams from the first base of gene). It has been confirmed that different promotor and

enhancer sequences are also involved in regulating the expression of ApoE gene [1]. For the protein expression ,mRNA of ApoE is 1163 bp in length [1][40]. Genes are specific for unique characters of any species [41]. Genetic makeup of any man or woman is known as its genotype [42].

In many species including human 99.9% of nucleotides are same. The unique behavior of any individual is due to 0.01% particular bases. This uniqueness in characters is because of differences which may be in coding or non-coding sequence [43]. This variance can be because of mutation or polymorphism. The mutation is a rare alternate in nucleotide sequence which may be point mutation or chromosomal aberration. Polymorphism is a variation in nucleotide which occurs often in any population. Polymorphism can be a single Nucleotide Polymorphism in which one nucleotide is modified. This nucleotide may also result in disastrous exchange in protein shape and its interaction or it could be harmless [44].

2.3 Site of Production of ApoE Protein

The primary site of manufacturing for ApoE is hepatic cells (75%). The second major supply of protein synthesis is the brain[45]. Even though major proportion of ApoE in cerebrospinal fluid is produced via astrocytes, in certain circumstances neurons also produce ApoE [46]. Macrophages and other cellular regions are also sites of ApoE manufacturing [1]. Maximum organs comprise large abundant amount of ApoE. The presence of ApoE has been confirmed in a variety of tissues in many different species [45][1][47].

The most important exception is in the case of the gut that is considered not to be concerned in ApoE manufacturing. The liver is a chief source of ApoE because of the availability of maximum mRNA present in it. Almost two-third to three-fourth of ApoE is produced with the help of the liver [45]. In each organ, numerous cells are responsible for ApoE production. This indicates the involvement of ApoE in lipid delivery and many different cellular activities. In liver, ApoE is produced by hepatic parenchyma cells [47].

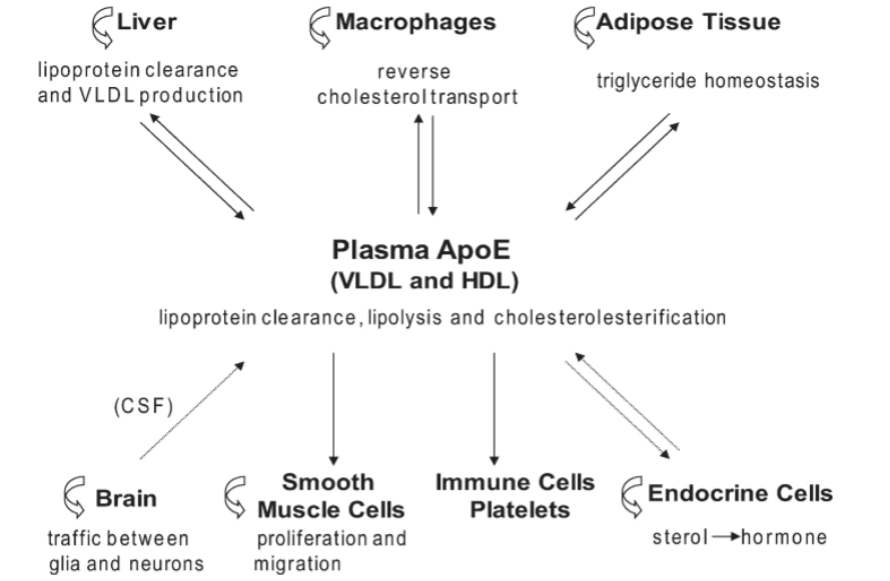

FIGURE 2.4: Different Sources of ApoE production[48]

It is considered that in the liver, ApoE is produced as factor of VLDL. However, production of ApoE as discoid particles containing phospholipid rather than a part of VLDL is also possible.[45] By loading mice peritoneal macrophages with cholesterol, synthesis and release of ApoE can be induced to a big degree [1].

In-vitro situations represent that newly synthesized proteins have 5%-10% ApoE. ApoE exists as ApoE phospholipid disks after its release from cells at the surface of phospholipids. Those disks have a density of approximately 1.08 gtml. Macrophages are also involved in liberating cholesterol and its production has no interaction with production of ApoE. Cholesterol binds to receptors in an extracellular fluid consisting of HDL and ApoE forms HDD-with apo-E on combining with HDL. Production and release of ApoE can be controlled by the state of activation of macrophages [1].

2.4 ApoE as a Multifunctional Protein

The function of ApoE involves lipid metabolism, regulating conversion, manufacturing, clearance of lipoproteins. Inside the liver, at some point of the assembly

of ApoE after its manufacturing, it binds to VLDL. Another source of becoming a member of ApoE with VLDL is the change of ApoE with other lipoproteins. As part of lipoproteins,ApoE facilitate the uptake of apoE-containing high-density lipoprotein (HDL) via the low-density lipoprotein receptor (LDLr), LDLr-associated protein 1 (LRP1), and heparin sulfate proteoglycans (HSPGs) into the liver. Inside the peripheral region, ApoE also helps in binding to the VLDL re-

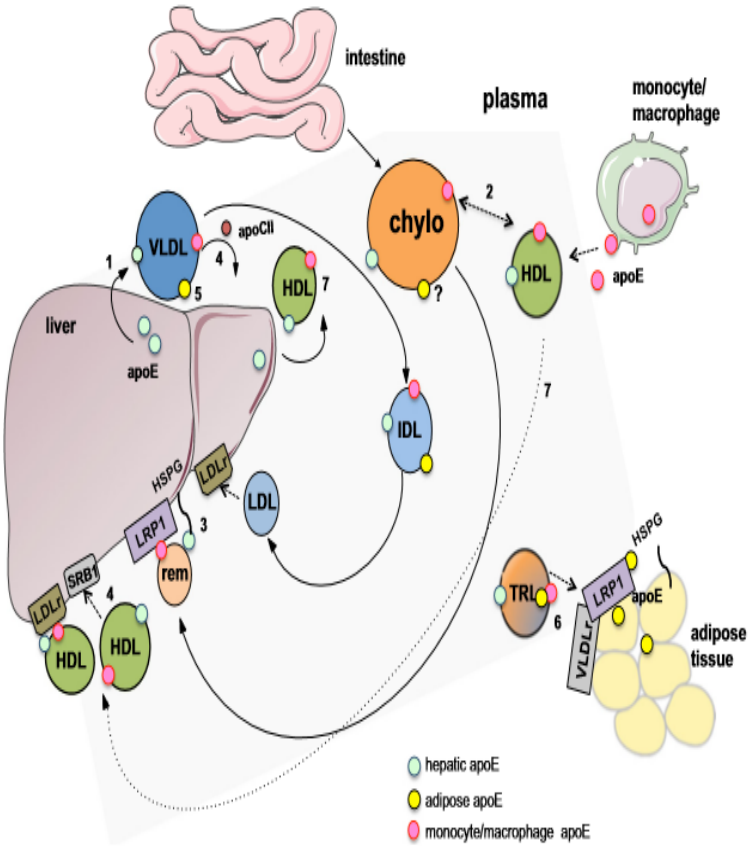

FIGURE 2.5: Functions of ApoE

ceptors on various tissues providing triglycerides for energy or storage as a part of VLDL. Different functions of ApoE include (1) Elimination of ApoCII from triglyceride rich lipoproteins. (2) Enhancing reverse cholestrol delivery in which cholestrol is transported to liver for removal through HDL [49]. Recent studies have suggested a strong affiliation between ApoE and CVD [50].

ApoE from macrophage assist to prevent atherosclerosis without affecting lipid level of plasma [51]. The anti-atherogenic impact of ApoE is due to anti-inflammatory anti-proliferative immune-modulatory properties [52][53][54]. ApoE from adipose

tissue plays a main role in affecting insulin sensitivity and causing adipose tissue irritation in mice [55][56].

2.5 Cholesterol Level in Blood and ApoE

Cholesterol levels vary in every organism. It has been investigated that a 50% variance in cholesterol level is due to genetic variability. ApoE gene locus is responsible for 16% of the variance in the level of low-density lipoprotein cholesterol. Organisms with E3/E2 have 20% lower, and those with E3/E4 have 10% higher amount of LDL cholesterol as compared to an individual with E3/E3 phenotype [57][58].

As cholesterol level is associated with atherosclerosis so it's been concluded that due to the involvement of ApoE in determining cholestrol level in blood, ApoE alleles also have role in coronary artery disorder [59].

2.6 Metabolic Pathways of ApoE

Apolipoprotein can bind with LDL receptors because it directs the uptake of lipid complexes or HDL with ApoE by using cells inside and nearby environment. Cells required cholesterol for various procedures which include membrane repair or cellular proliferation.

Within the macrophages, ApoE behaves like "autocrine". ApoE secreted by way of macrophages attaches with lipids and circulate them to macrophages for its storage. ApoE transports lipids and dietary fibres to the liver and peripheral tissues.

Gut synthesize chylomicrons due to uptake of dietary fat and cholesterol, which enter in mesentric and thoracic duct lymph [60]. Lipoprotein lipase is an enzyme concerned with the hydrolysis of triglycerides of chylomicron and the fatty acids produced in the result are stored in adipocytes as triglyceride droplets. After

hydrolysis, chylomicrons contain a massive quantity of cholesterol and known as chylomicrons remnants these chylomicron remnants are absorbed by the liver from plasma wherein cholesterol is either required for membrane lipoprotein synthesis or it is far secreted as bile. It has a role in facilitating the uptake of remnants by the liver. The receptor which is accountable for the uptake of remnants is located on hepatocytes and is not like apo-B,E (LDL) receptor. This specific receptor is named as chylomicron receptor or ApoE receptor [61]. The prevalence of this receptor is postulated because of looking at the low degree of chylomicron remnants in plasma of patients with absent or defective LDL receptors [1]. It does not display that LDL receptor has no role for chylomicron remnants uptake however it shows that ApoE receptors also are involved in this process other than LDL receptors [62]. So, it is true to consider that uptake of chylomicrons remnants by using liver in a multistep manner and although the actual mechanism is not regarded however its confirmed that ApoE favours the uptake of chylomicron remnants.

ApoE also plays an effective role in the transfer of lipids from the liver to peripheral tissues. Very low-density lipoproteins are triglyceride rich lipoproteins that might be released from hepatocytes containing Apo E and ApoB 100. Within the plasma lipoprotein lipase act on this and convert it into VLDL remnants like chylomicron remnants.Fatty acids are removed from VLDL and render them cholesterol rich particles. These VLDL remnants are processed through a chain of reactions and finally transformed to LDL by passing from IDL state. During this, a fraction of remnants is also cleared from plasma by using LDL receptors. ApoE additionally has a role in the uptake of VLDL and IDL when lipoproteins transforms into LDL and lose their ApoE then Apo B 100 is helpful for its binding with LDL receptors [1].

ApoE is also involved in a process called reverse cholesterol transport which was postulated many years in the past and in keeping with that pathway HDL is a critical source of transport. In the interstitial fluid, cholesterol is released from cholesterol loaded cells together with macrophages to cholesterol acceptors. One of the cholesterol acceptors is HDL, specifically the phospholipid-rich, nonapo-E-containing HDL. By means cholesterol availability HDL become to be cholesterol

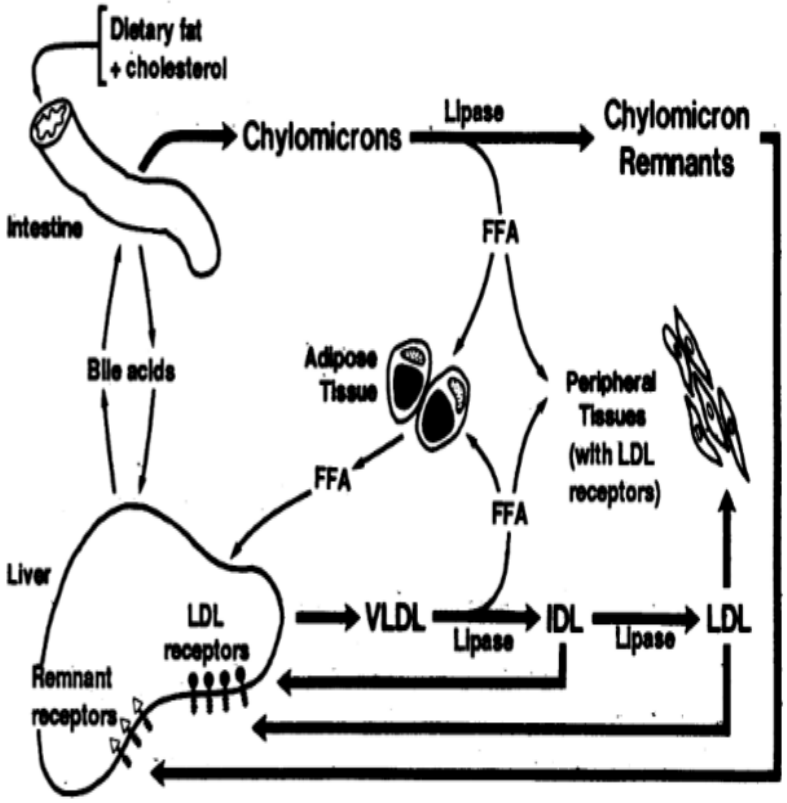

FIGURE 2.6: Metabolic Pathway of ApoE

rich and it requires ApoE. ApoE is in interstitial fluid as released from a variety of cellular types as macrophages, easy muscle cells, and others.

ApoE binds with HDL when they have become cholesterol rich. It has been proved by using invitro research that ApoE in interstitial fluid allows the uptake of cholesterol by HDL. When extra ApoE is provided to a system setup of an experiment having cholesterol and HDL with an enzyme then it shows the formation of very cholesterol enriched HDL with double diameter.

The cholesterol esters become organized in concentric layers inside the center of particle, and available cholesterol content material on the surface of HDL particles will be increased which favors its binding with LDL receptors.In the absence of ApoE HDL can not bind with LDL receptors. So, cholesterol enriched HDL with ApoE in the interstitial fluid can transfer cholesterol from peripheral tissues to liver in which it can bind with LDL receptors.Presence of such cholesterol enriched HDL with Apo E has been confirmed in lots of animals. In humans HDL with

ApoE are present in low amount and reverse cholesterol transport includes an extra mechanism in which cholesterol ester transfer protein is found to be involved in transport of cholesterol ester from one lipoprotein to another.It works through trasport of HDL cholesterol ester to low density lipoproteins including VXLDL, IDL, or LDL, which then are taken up by the liver. However additionally in human beings a few cholesterol is transported without delay to liver utilizing reverse cholesterol mechanism [1].

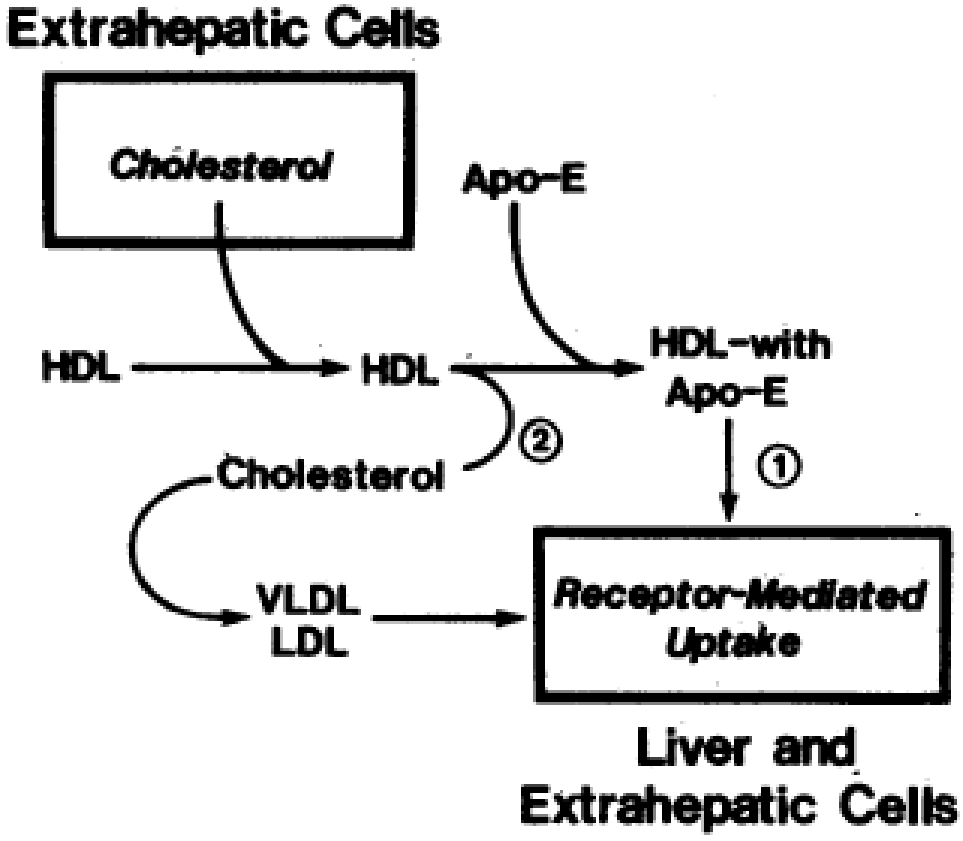

FIGURE 2.7: Mechanism of reverse cholesterol transport

2.7 ApoE in Lipid Redistribution

It is possible to study the role of ApoE with HDL or lipid complexes in the transport of cholesterol from one organ to its local area or to different organ. So it can be said that ApoE plays a role in the redistribution of ApoE from organ with high level of ApoE to organ with lower ApoE level. To facilitate the uptake of

ApoE, cells on surface of organs must possess increase in LDL receptors. Role of ApoE was studied in the transport of cholesterol among the cells of damaged and regenerating nerves of peripheral system.This role is presented in the figure 2.7. In this model it was observed that the level of ApoE production increases almost 100-200 fold in case of any injury or cut in sciatic nerve as compared to healthy nerve. The level of ApoE in extracellular regions is 5% of total protein in the nerve region which is in the process of regeneration.It takes almost 7 to 10 days for increase in ApoE level after injury and then level gradually returns back to its normal level by 8 weeks. ApoE in this case is mostly produced by macrophages. Soon after injury, macrophages in the sciatic nerve start producing ApoE. Soon after the injury, monocytes enter at the site of injury due to inflammatory action and becomes macrophages and start secreting ApoE.Then major lipids start accumulating in the damaged area in schwwan cell and macrophages.

2.8 ApoE in coronary Artery Disease

Coronary artery Disease (CAD) is a multifactorial disease this is considered to result from an interaction among genetic heritage and environmental elements consisting of weight-reduction plan, smoking,and exercise. It is also related to conventional hazard factors, inclusive of high blood pressure, diabetes mellitus, and hypercholesterolemia. However, in some people, CAD is not related to such threat factors,suggesting that different genetic elements make a contribution to a predisposition to coronary atherosclerosis and its thrombotic issues. In general, individuals with high blood pressure,diabetes mellitus, and hypercholesterolemia and people without any of those elements are considered at excessive and low threat,respectively, for development of CAD. It's far for this reason crucial to perceive genes that confer susceptibility to CAD in these high- and low risk people independently. Genetic epidemiologic studies have recommended that specific genetic changes, which includes polymorphisms within the genes encoding platelet glycoprotein IIIa, methylene tetrahydrofolate reductase, and plasminogen-activator inhibitor-1 , are related to occurrence of CAD in excessive- or low-chance

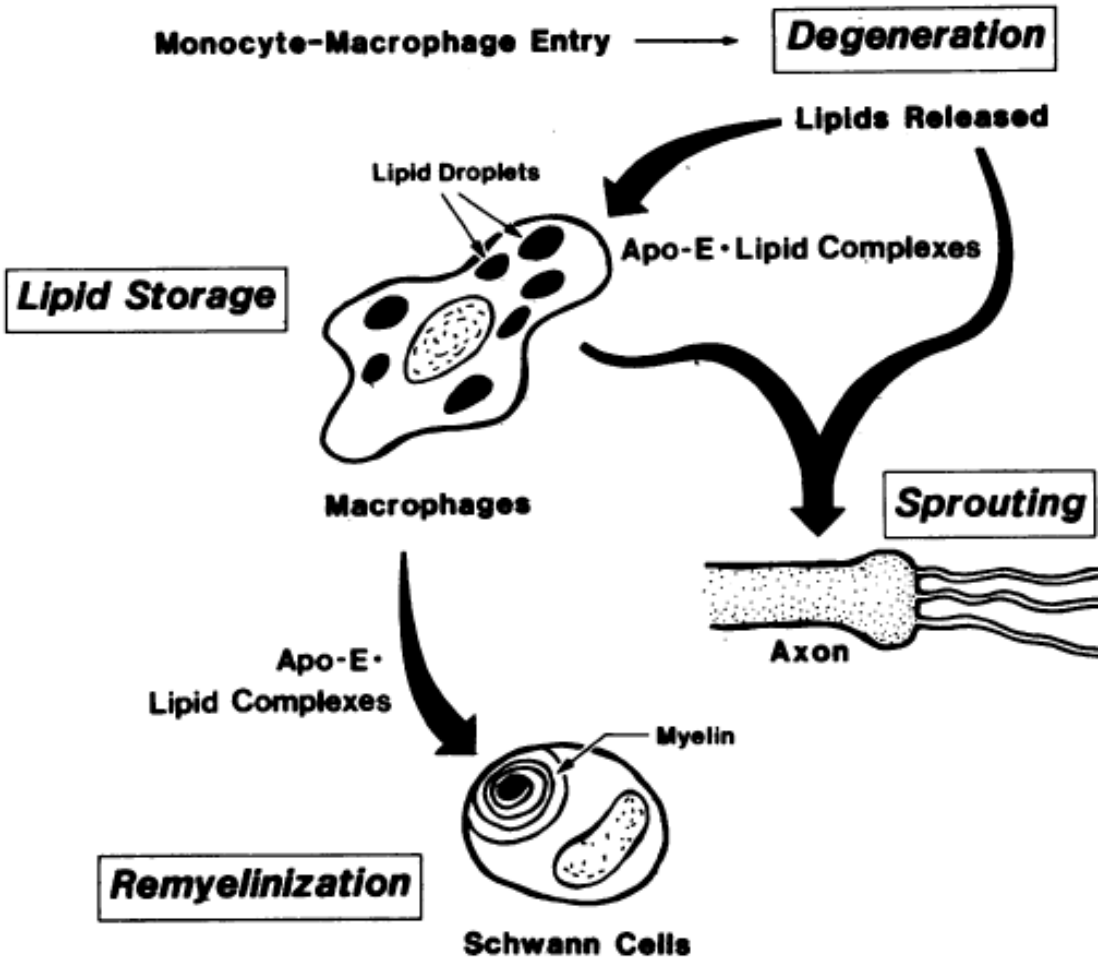

FIGURE 2.8: Role of ApoE in reverse lipid transport

individuals. But, the genes that make a contribution to genetic susceptibility to CAD in people with predominant conventional threat factors high blood pressure, diabetes mellitus, and hypercholesterolemia or in those with none of those elements stay to be recognized. Similarly, due to ethnic divergence of gene polymorphisms, it is important to study polymorphisms associated with CAD in low or excessive threat people of each ethnic group. Apolipoprotein E (ApoE) is a protein this is incorporated into serum lipoproteins and directs their catabolism through binding to receptors. ApoE modulates the catabolism of triglyceride-depleted remnants of chylomicrons (CM) and very low-density lipoprotein (VLDL). The frequencies of the E2 (112cys and 158cys), E3 (112cys and 158arg), and E4 (112arg and 158arg) alleles are noticeably regular in grownup Caucasians (8%, 78%,and 14%, respectively). However, the frequencies of those alleles in other populations aren't equal.

Allelic differences on the apo E gene locus determine approximately 16% of the genetic variance in the concentrations of LDL-C. Consequently individuals with the genotypes E3/E2 and E4/E2 have 20% lower and people with genotype E4/E3 have 10% better stages of LDL-c as compared with subjects with genotype E3/E3.

The occurrence of CAD is increasing in India, specially in the younger population. Subsequently, it's miles critical to determine susceptibility factors for CAD in Indians. No statistics are available for the connection between apolipoprotein gene polymorphism and their plasma degrees or different CAD risk elements in inhabitants of Asian Indian population in eastern a part of India. We consequently hypothesized that genetic element of these apolipoprotein gene (apo AI, apo B and apo E) polymorphisms are vital determinants of CAD on this population [35, 36].

2.9 ApoE and Alzheimer's Disease

Alzheimer's disease (AD), an age-associated neurodegenerative problem, is the main type of dementia and a main reason of death in the old people. Its incidence is increasing readily with the getting old populace throughout the world. However,mechanism responsible for causing this disease stays doubtful. Over the last few years, numerous genetic causes for late-onset Alzheimer's disease (LOAD) were identified, which includes non-coding variations with low penetrance (odds ratios = 1.05–1.30).

Especially, the APOE locus tagged by using coding variation APOE-ε4, is unequivocally the maximum significant risk factor for AD. At the same time as different risk factors have also been indicated in this location. The ε4 which is one of the common isoforms of the APOE gene is a well-reputed risk aspect for Alzheimer's disease, dementia with Lewy bodies in addition to vascular dementia, and has also been determined to be related with cognitive functioning with out dementia. Further to the ε4 main isoform, the functional polymorphism rs405509 inside the which is in the region of APOE gene promoter has been discovered to be associated with severe risk for AD in addition to nonpathological cognitive aging.

However, research in this polymorphism has no longer been discussed. Some researchers have discovered better AD risk related to the mostly occuring allele and others with the minor allele. Moreover, some researchers have declared risk factor which are different for patients of different age group. It has additionally been not clear that either the promoter polymorphism's effects in the promoter region are truly unbiased of or because of linkage disequilibrium with APOE common isoforms. It's been advised that the APOE locus is related to nonpathological cognitive ageing and dementia risk through different mechanisms. However, there have up to now been no research to find relation of APOE essential isoforms and rs405509 promoter polymorphisms with both results in a single experimental population.It has been further reported that at the same time as the APOE fundamental isoforms had been now not associated with cognitive potential in aged men from the Helsinki beginning Cohort observe (HBCS), the quantity of rare alleles in rs405509 and in every other functional polymorphism, rs440446 in the APOE gene intron-1,depicted better preserved cognitive potential unbiased of the APOE major isoforms.

2.10 ApoE and HCV infection

Hepatitis C virus (HCV) disease is a considerable health related problem which is a cause of continual liver sicknesses worldwide.According to the world health Organization(WHO), 71 million human beings are chronically affected, and 399,000 deaths every year are associated with HCV infection. Estimates are that up to 90% of the infected individuals are unaware of their source of contamination. In approximately 25-30 years, persistent HCV contamination may additionally steadily cause a vast spectrum of clinical consequences which includes fibrosis, cirrhosis, and in a few cases, hepatocellular carcinoma [3]. However, a few patients (20%-40%) might also clear up an acute infection by means of self-spontaneous clearance of the virus, evidenced through fine anti-HCV antibodies and negative viral RNA in the serum. This chance is variable because of an aggregate of the immunologic, metabolic, and genetic elements of the host.specially, plasmatic degrees of total

cholesterol (TC) and low-density lipoprotein ldl cholesterol (LDL-c) were stated as predictors of the reaction to interferon therapy for the duration of HCV infection.

Likewise, the Apolipoprotein E (APOE) gene encoding the glycoprotein element of the low-density lipoprotein has also been implicated within the final results of HCV infection and related comorbiditie. HCV binds to the ApoE ligand getting into the hepatocyte via the low-density lipoprotein receptor (LDLR). functional polymorphisms rs429358 and rs7412 in the APOE gene lead to a few common alleles, ε32, ε3three and ε3four, encoding the corresponding essential isoforms, ApoE -E2, -E3 and -E4. ApoE3 is the wild-kind isoform with a natural affinity for the LDLR, even as ApoE2 and ApoE4 present adverse binding competencies. ApoE2 isoform has a considerably decreased attachment to the LDLR.

Conversely, the ApoE4 isoform confers high binding to LDLR as compared to ApoE2 and ApoE3. Those relative binding abilities are steady with findings that endorse a protective effect of APOE ε3four in the development of liver damage as revealed by means of histopathological analysis[eleven], whereas APOE ε3 has been associated with the powerful infection. There may be growing proof of the occurrence of dyslipidemia in HCV-infected patients. consequently, variation in body weight may also have a meaningful effect at the management of those patients. presently, Mexico and the US are experiencing a large person weight related health problems.

In Mexico, 72.5% of the grown up population has the problem of overweight or obesity. This increase in Bio mass index (BMI) is associated with the development of several metabolic abnormalities such as dyslipidemias together with, hypercholesterolemia (HChol), that's one of the eight maximum crucial hazard factors of mortality in Mexico. Each obesity and dyslipidemia are related to environmental chance elements together with diet. Currently, it has been described that the nutritional pattern of the overall Mexican populace and HCVinfected patients favours the development of lipid abnormalities. On the opposite side, the APOE ε34 allele that is related to HChol has a heterogeneous occurrence throughout the country population starting from 0-20.3%. However, the relationship between

APOE alleles and lipid metabolism, in addition to its potential implication in HCV infection most of the Mexican populace is currently unknown. West Mexico is a place characterized by way of a genetically admixed populace with Amerindian, Europeans, and much less significantly African ancestries. Given the variability of APOE alleles found with the aid of ethnicity, it's miles attainable that variations in the genetic and environmental elements of the Mexican population might also influence the relationship among APOE, lipid abnormalities and outcome of HCV infection [39].

2.11 ApoE and Renal Infection

Chronic kidney disease (CKD) is one of the current leading public fitness issues due to growing frequency,complications as well as high mortality attributable to expanded atherosclerosis. the main cause of dying in this population is cardiovascular disease (CVD) related with dyslipidaemia, which is discovered at early degrees of renal failure and related to the extent of glomerular filtration rate declining.

Hypertriglyceridemia,accumulation of intact or in part metabolized apolipoprotein B (APOB)-containing lipoproteins (very low-density lipoprotein (VLDL), intermediate density lipoprotein (IDL), low density lipoprotein (LDL)), and reduced attention of high density lipoprotein (HDL) are properly reported lipid disturbances in CKD.other studies have also shown that the amount and distribution of apolipoproteins - which play a vital position in lipid metabolism - also are disturbed in CKD.

Apolipoprotein E (APOE), a 34 kDa glycoprotein, is produced in particular in the liver however its local manufacturing has been documented inside the mind, kidneys, spleen, adrenals and macrophages. It is a component of all groups of lipoproteins except small dense LDL. It circulates among HDL and APOB-containing lipoproteins in plasma . APOE acts as a ligand for receptor-mediated clearance of chylomicrons (CM) and VLDL remnants from the movement to the liver and takes

element in reverse ldl cholesterol transport (RCT) as a component of HD. The defensive position of APOE in atherosclerosis development was first established by experiments on animal models. In further studies it was proven that APOE is an anti-atherogenic protein additionally in human beings. It was also studies that cholesterol-loaded macrophages validated better APOE gene expression; APOE deficiency in these cells decreased ldl cholesterol efflux and led to atherosclerosis plaques formation.

Due to the variations in APOE isoforms structure, they own numerous binding ability to receptors in addition to lipoprotein-binding options that effects in a diverse way lipoproteins metabolism. APOE3 and APOE4 isoforms bind to the LDL receptor with similar amazing affinity, whereas APOE2 is defective in interplay with the receptor - almost 2% of regular activity.

APOE2 and APOE3 isoforms preferentially bind to small HDL, at the same time as APOE4 to massive triglyceride(TG)-rich VLDL particles. APOE3 is the most commonly occuring isoform (around 77% of the Caucasian population) and does no longer impair lipoprotein metabolism. APOE is present in 15% of the population and can lead to increased LDL-C concentration. APOE2 is the rarely occurring isoform and is probably related with hypertriglyceridemia.these disturbances in various ways make a contribution to increase in risk for atherosclerosis.

The APOE gene polymorphism additionally influences APOE level in serum; the highest ranges had been measured in $\varepsilon 32$ allele carriers, the lowest in individual with heterozygous for $\varepsilon 34$. Even though APOE has anti-atherogenic capabilities, there are ambiguous facts that confirm the link between APOE level and CVD threat. Surprisingly, it has been shown in lots of research that CVD patients had increased APOE level. However, researchers have highlighted that APOE distribution among lipoproteins has a critical means and total APOE amount in serum is not the perfect parameter for the assessment of CVD chance. It's been validated that accelerated APOE level within the CVD populace resulted from higher APOE content material in APOB-containing lipoproteins and that component may be a proper marker for this purpose. The crucial position of APOE in lipid metabolism

and CVD improvement in CKD patients has also been emphasized by many researchers. The polymorphism of the APOE gene has an influence on APOE level in individuals as well as lipids and lipoproteins disturbances.Wang Y et al. proved that hemodialysed ε3four allele carriers have higher overall cholesterol (TC), TG, and LDL-C.

ε32 allele carriers exhibit accumulation of ldl cholesterol-rich VLDL. but, there may be incomplete information regarding the relationship between APOE gene polymorphism, APOE level, its distribution amongst lipoproteins and the degree of kidney disorder, specifically at early ranges of CKD. those findings ought to shed new light on the pathomechanism of lipid-associated problems and increased atherosclerosis in CKD patients.

2.12 ApoE and Head and Neck Cancer

Head and neck cancer is characterized via mobile malignancy within the higher aerodigestive place, such as paranasal sinus, nasal hollow space, oral cavity, pharynx, and larynx. Global, the occurrence of head and neck cancers is set 550,000 instances yearly with a mortality chances about 300,000 instances in 12 months . In 2008, The International Agency for research (IARC) conducted the GLOBOCAN assignment, with an predicted result 262,600 new cases of esophageal cancer with 223,000 mortality charge and 107, seven hundred new cases of oral cancers with 61,200 mortality rate of fellows in growing countries. But, new instances of esophageal most cancers in girls showed an expected result reached about 137,900 with morbidity rate as much as a 115,900.

In growing countries (2008), the incidence of oral cancers reached 4.6% in men and 5.7% in girls. The foremost dangers component for head and neck most cancers is tobacco and alcohol consumption. Further, there are different threat components that could affect the occurrence of head and neck cancers, one of them is genetic element. Genetic composition has a function in susceptibility of disease because of gene polymorphism. Some researchers have proven that head

and neck cancers are associated with certain gene polymorphisms. Gene have a position in encoding human enzyme to make a contribution in metabolism of poisonous materials and also impacts human susceptibility from dangerous impact of carcinogenic materials.

One of the genes that contribute is Apolipoprotein E (APOE). The feature of this gene is to make protein (apolipoprotein E) and combine with lipid to form lipoprotein. those lipoproteins will deliver lipid and cholesterol via blood vessel. Further, APOE also have various features which includes tissue repair, immune response, regulation of synthesis, metabolism of bile acid, mobile growth, cellular differentiation, metastasis, and angiogenesis.

APOE genes have three allele with 6 genotype possibilities. Every genotype has the potential to bind special receptors resulting in extraordinary low-density lipoprotein (LDL) level. Studies indicates, high ranges of LDL has inverse relation with the hazard of most cancers possessed which means that the higher LDL level, the decrease the danger most cancers. But, other studies have demonstrated that low LDL stages which are the result of head and neck cancer now not a reason.

2.13 ApoE and its Relation with Longetivity

Longevity is typically taken into consideration because the result of interactions between environmental and genetic factors. Some of the environmental elements, life-style and the nutritional value seem to play an important function within the prevalence of many age-related diseases.As an example, high stages of adiposity, people with body mass index (BMI) more than 30 kg/m2,were linked to strange glucose metabolism and multiplied incidence of cardiovascular sicknesses(CVDs), cancer and neurodegenerative sicknesses, with therefore increases in mortality rate. Furthermore, due to the fact CVDs are the primary reasons of loss of life globally, polymorphisms in genes involved in those types of illnesses, in addition to those involved in vascular dysfunction and weight problems, contribute much for longitivity. Apolipoprotein E (APOE) gene is one of the factors constantly associated

with human longitivity. The glycoprotein encoded by way of this gene performs a fundamental role inside the lipid metabolism, mediating lipoprotein binding to the LDL and lipoprotein remnant receptors. Any structural change on this protein may cause a decrease in its capability to bind to the receptors, which as a result cause an increase in blood cholesterol level and, as a result, to extended risk of atherosclerosis. The E4 allele is considered to be the commonly occuring APOE allele in all human populations and it's responsible for increase in total cholesterol levels and increase risk of metabolic and neurodegenerative illnesses,whereas E2 and E3 alleles had been associated with decrease cholesterol levels and with a protective function in opposition to Alzheimer's disorder. Any other gene related to CVDs is the ACE gene, that encodes an enzyme that catalyzes the conversion of angiotensin I to angiotensin II, a strong vasoconstrictor [41, 42].

2.14 Protein Protein Interaction (PPI)

It's been studied that genotype of any individual has the power to decide its phenotype. Genotype expression includes the production of proteins and these proteins do no function individually. One protein has interaction with other protein or also with other biomolecules such as DNA and RNA. By those interactive community proteins functions in distinctive pathways [63]. Studies of PPI may be helpful in the subject of medication. The PPI involves structural as well as functional interaction.[64]The statistics of PPI may be used to generate protein interaction pathways.[65][66].Even though data regarding PPI is not complete [67]. However the available facts can be used successfully through the use of superior technical strategies.[68] Many disorders involve the mutation in PPI [69]. So, to take a look at any ailment at molecular level protein interplay observe may be beneficial [70]. There are common methods for observing the interplay of proteins. In conventional method, interplay of one protein with different protein is studied and under the superior technique the interaction of one protein with more than one proteins is studied and the resultant facts is used to construct a community of protein interaction [71].

2.15 Pathway Analysis

The relationship of structure and function has no longer been investigated completely in complicated cell networks. Systematic biology tries to explain this relation using aggregates of each experimental and theoretical procedures. Many theories have been proposed for explaining protein networks. Each of those theories has its benefits and shortcomings. Dynamic mathematical modeling of huge-scale networks has the issue of the availability of mechanistic detail and kinetic parameters. In comparison, structure-oriented analyses best need only network topology, which is well known and without difficulty to be had in lots of instances [72].

In recent years, definitions of biochemical pathways are primarily based on their network therefore, stability use of an entire network of biochemical reactions is required. Strategies, elementary modes, and extreme pathways have formulated novel hypothesis addressing biochemical network function. An interaction exists among those two processes [73].

Genetic manipulation of cells also can enhance the yield and talent of metabolic response. Metabolic networks can be reconstructed or a few pathways can be deregulated. This has resulted in the metabolic engineering idea inside biotechnology which ends up in novel products, favored products, or controlling wastage of strength on the production of pointless compounds by using adjustments in existing pathways. The term 'metabolic engineering is used for analyzing metabolic states and predicting useful changes in them through the usage of appropriate tools [74].

2.16 Approaches of Pathway Analysis

Although a large quantity of data has been produced per test, this also calls for a way to translate these records into an understandable form. Most systems and analysis techniques bring about many lists of differentially expressed genes as their output. Most researchers need to face the difficulty of translating this sort

of big data for the know-how of biological system and also studying these records with recognize to complete organism considering it a complicated community. In 2002, an approach using the Gene Ontology (GO) become proposed to address this difficulty. This method requires the data of differentially expressed genes and predict go categories (e.g. biological processes, and many others.) by the usage of statistical evaluation. This method works by the comparison of a variety of differentially expressed genes with the range of genes predicted to be determined accidentally. The study is taken into consideration to be useful if the determined variety is appreciably distinct from predicted just by chance. Presently, there are over 20 tools that use this over-representation approach (ORA). Even though this approach is broadly followed, it also has some of the limitations that are associated with the sort, and structure of the annotations available [75].

A biological pathway that arranges records concerning genes and proteins into an interpretable set of altered mechanisms and combines records from many experiments has a beneficial role in understanding the underlying mechanism of various diseases, enhancing the impact of clinical treatment, and discovering drug objectives and biomarkers. In addition, many pathway evaluation models have been designed that depict the laws of life activities and employ network topology statistics. Although most of these strategies are based on network topology facts, they most effectively use the topological model of the pathway and don't think about the facts of pathway outside genes in living networks; consequently, they do not fully process the pathway records [76].

Stoichiometric network analysis (SNA) of biochemical reaction systems is an essential approach for the know-how of the functionality of metabolic networks reconstructed by way of genomic and biochemical records. SNA is depending on the mass balances of metabolites and the belief of metabolic pseudo-steady state which leads to the capacity of a homogeneous system of linear equations set up with the aid of the stoichiometric matrix N and the vector r of internet reaction fees: (0 = Nr) while metabolic flux analysis focuses on single flux distributions, Metabolic Pathway evaluation research the whole set of admissible flux distributions [77].

Chapter 3

Materials and Methods

Due to the involvement of various alleles of Apo E in different diseases, structure and function were predicted by using a bioinformatics tool. Interactors were also identified to check the various pathways affected by Apo E variation.

3.1 Determination of Deleterious SNPs

To detect the effect of different SNPs, following steps were performed.

3.1.1 Sequence Retrieval

The sequences of nucleotides for apo E were collected from NCBI (National Centre for Biotechnology Information). NCBI is an online database with information regarding nucleotides sequences and amino acid sequences. Total 2001 SNPs have been reported.

Among all these 25 SNPs were selected for which protein sequences have been determined. The sequence of protein for following SNPs was also retrieved from UniProt [15]. Uniprot is a database of information about protein structure and function.

3.1.2 Identification of Deleterious SNPs

A single SNP can affect the phenotype of an organism. It can cause disease sometimes [16]. Among the 25 SNPs, deleterious SNPs were selected by the sift server. SIFT is an algorithm which uses sequence homology based approach to check aminoacids substitutions. It also uses information from evolutionary conservation of amino acids. The results obtained from the sift server were crosschecked by Polyphen 2. Deleterious effect was confirmed by using Polyphen 2.

Polyphen 2 is a software which detects the effects of substitutions of amino acids on the stability and function of proteins. Polyphen 2 features involve a high quality multiple protein sequence alignment pipeline and a prediction method using machine learning classification. The specificity value of each SNP is used as an indication of the damaging effect of that variant. 1% of specificity confirm the damage. It helps in the prediction of the effect of the change on the stability and function of the resultant protein [17].

3.1.3 Gene Ontology

For this purpose, GO Terms for damaging SNPs were collected by GO Term Finder EBI. By using SNP & GO, structural and functional properties were determined [18]. For the confirmation of deleterious SNPs, SNP & GO uses functional annotation [19]. Mutpred also confirms the deleterious effect of variants [20].

3.2 Protein Stability

The single amino acid has the potential to change the stability of the protein [83]. The stability of the variant protein was checked by using I mutant 2.0. This tool does not give an exact value to which extent any protein is stable.RI is calculated in this tool which is an indication of the extent of stability of any mutated protein [84].

3.2.1 Evolutionary Conservation

Evolutionary information was checked by using the Consurfserver. Evolutionary data helps in the understanding of protein structure and function relationship [85]. Some amino acid changes were not passed along the generations while others evolved slowly having a considerable role in evolution [86].

3.3 Structure and Alignment of Mutated Proteins

Structures of mutated proteins were predicted and validated and then these predicted structures were aligned with wild-type protein structure by following steps.

3.3.1 Structure Prediction and Validation

The structure of the protein was built up by phyre 2. That structure was cross checked by I-tasser [87]. I- trasser is considered to be the best of all available servers for 3 D structural prediction. This tool uses information for the prediction of the structure and functions of the mutated protein [88]. The predicted model was validated by save v5 [89]. The structures predicted by I trasser were similar to those predicted by Phyre 2.

3.3.2 Structural Alignment

Alignment of the predicted model was done by using TM align. Wild type model was retrieved from PDB server. The alignment was done to determine the variation due to SNPs [90].

Structural alignment by TM align has higher accuracy as compared to all most often used methods. This tool compares the predicted structure with all available protein structures in the protein data bank [91].

3.3.3 Post Translational Modification

Post-translational modification involves structural changes. It also involves the methylation of protein.

These modifications were predicted by pssme and then crosschecked by bpb-ppms. Furthermore type of methylation was checked by the web server [92].

3.4 Interactors Identifications and Network Analysis

Interactors were identified to build the interaction network of ApoE. These interactors help detect the role of ApoE in pathways of different diseases.

3.4.1 Interactors Identification

Multiple interactors for the Apo E gene were identified by funcoup. This tool uses the functional analysis of proteins. Due to this, all the proteins having an association with apo E were detected [93].

3.4.2 Construction of Apo E Interaction Network

Genemania is a server for network analysis. It predicts the related genes involved with one gene and also describes the pathways in a system involving those genes. Interaction Data from this server was retrieved showing source and target genes. Interaction data was retrieved by genemania.

This network helps in the detection of nodes and edges. These nodes and edges were collected in the form of excel sheet [94]. Networks were checked in Gephi tool. This tool is responsible for the prediction of the degree of interaction [95].

3.4.3 Enrichment Analysis

Enrichment analysis is responsible for predicting highly correlated networks. All other pathways were excluded from this analysis. For this analysis, The Enrichment tool was used [96].

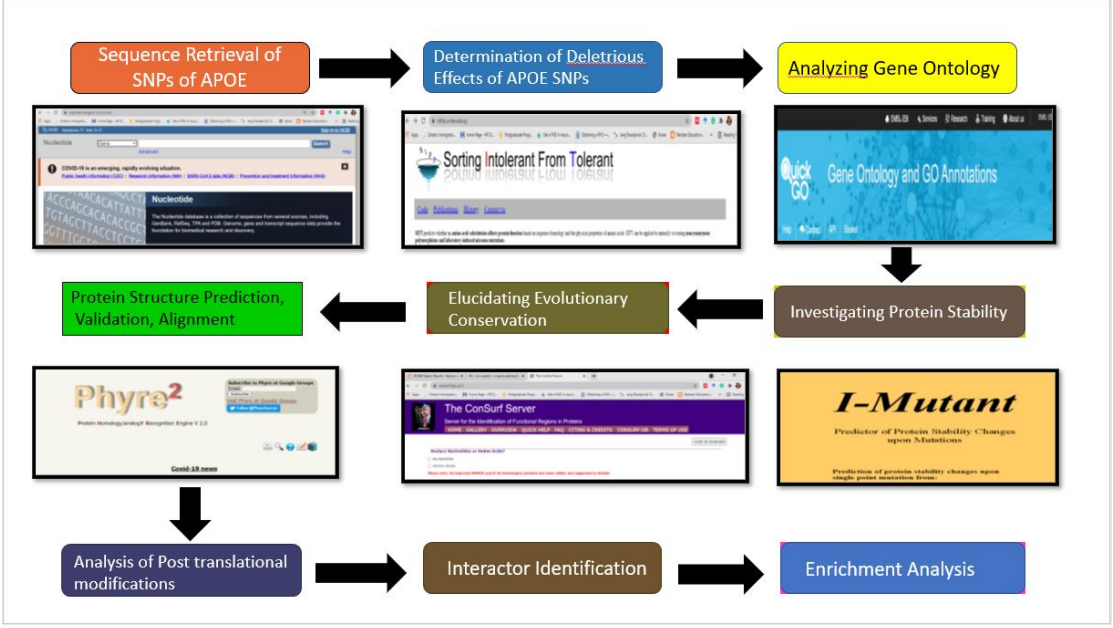

FIGURE 3.1: summary of methodological steps involved in determination and impact analysis of deleterious SNPs of APOE

Chapter 4

Results and Analysis

4.1 Sequence Retrieval

Sequences of nucleotide and amino acids were selected. Redundant sequences were excluded. The total SNPs were 2001, among those 25 SNPs were selected. These 25 selected sequences were for different variants produced by possible mutations at different positions. Rs ids for these sequences are as follows: rs7412, rs429358, rs769452, rs11083750, rs28931576, rs28931577, rs28931578, rs28931579 rs121918393 , rs121918394, rs121918397, rs199768005, rs267606661, rs267606662, rs1180612218, rs769455. The rs ID of any SNP is the reference SNP cluster ID. These rs ID's involve the following change of nucleotides at corresponding positions.

Table 4.1: Following table shows us rs ID & Variation.

rs ID	**Variation**
rs7412	C >T
rs429358	T >C
rs769452	T >A,C
rs11083750	C >A,G,T
Rs28931576	A >G,T

Table 4.1 continued from previous page

rs ID	Variation
rs28931577	G >A
rs28931578	G >A,C,T
rs28931579	A >C
rs121918393	C >A,T
rs121918394	A >C,G
rs121918397	G >A,C
rs199768005	T >A
rs267606661	C >G,T
rs267606662	G >C
rs1180612218	C >A
rs769455	C >T

4.2 Identification of Deleterious SNPs

Either any variant is tolerable or deleterious was checked by the sift server. The deleterious SNPs were crosschecked by polyphen 2. Tolerated SNPs were excluded from research to predict the effect of damaging SNPs. Deleterious SNPs are presented in table 4.2. The input was in the form of rs ids of 25 variants. Among those, the deleterious effects of 5 SNPs were confirmed. In polyphen 2 all these SNPs were detected as probably damaging except one which is possibly damaging. Probably damaging SNPs are considered more damaging as compared to possibly damaging SNPs. Sift Score is an indication of chances of finding a new amino acid at any position in the structure of protein. Its value ranges from 1-5. Proteins with sift score of 0-0.5 are considered to be having damaging or deleterious effects. All these SNPs have a sift score of <0.05 in the sift server which an indication of the deleterious effect of that SNP. Sift Median is the median value for the probability of that SNP in protein.

Table 4.2: Damaging SNPs of APOE retrieved through Sift Server and Polyphen 2

SNP	**rs7412**	**rs769452**	**rs28931577**	**rs28931578**	**rs769455**
COORDINATE	45412079	45411110	45411902	45412008	45412040
REF ALLELE	C	T	G	G	C
ALT ALLELE	T	C	A	A	T
AMINO ACID CHANGE	RI76C	L46P	A143T	R152Q	R163C
TRANSCRIPT ID	ENST00000446996	ENST00000252486	ENST00000434152	ENST00000252486	ENST00000252486
PROTEIN ID	ENSP00000413135	ENSP00000252486	ENSP00000413653	ENSP00000252486	ENSP00000252486
SIFT SCORE	0.001	0.075	0.03	0.029	0.08
SIFT MEDIAN	3.02	2.95	3.04	2.84	2.75
NO OF SEQS AT POSITION	24	18	25	26	27
SIFT PREDICTION	Deleterious	Tolerated	Deleterious	Deleterious	Deleterious
AVG ALLELE FREQ	0.075	0.001			0.007
EAS ALLELE FREQ	0.1	0			0
AMR ALLELE FREQ	0.048	0			0.006
AFR ALLELE FREQ	0.103	0			0.025
EUR ALLELE FREQ	0.063	0.004			0
SAS ALLELE FREQ	0.044	0			0

4.3 Gene Ontology

Deleterious effects of SNPs were confirmed by SNP & Go. For this tool, Go terms were collected for selected 5 damaging SNPs by using Go Term finder EBI. GO term is related to the functions of different biomolecules.

Table 4.3: GO Terms for selected SNPs

Sr. No.	rs ID	Uniprot ID	Go Term
1	rs7412	P02649	GO:0001540
2	rs769452	A0A0S2Z3D5	GO:0060228 GO:0005615 GO:0006874
3	rs28931577	P02649	GO:0042803 GO:0061771 GO:1990777
4	rs28931578	A0A0S2Z3D5	GO:0071813 GO:0010877 GO:0034362
5	rs769455	P02649	GO:0005198 GO:0043254 GO:0005576

Its value changes due to a change in the function of molecules. GO term value is helpful in GO term enrichment analysis. Uniport ID and Go Terms of different variants were given in table 4.3. SNP & Go is considered to be the best of all servers for prediction of the effect of the mutation on protein due to its user-friendly nature.Protein Sequence,Go Term, and mutated amino acids names were given as input and the results were about the RI and probability value. RI is reliability Index value which is an important indication of the strength of deleterious effects of various SNPs. Results of SNP & Go were presented in **Table 4.4**.

TABLE 4.4: RI value of Diseased SNPs

Rs Id	Mutation	Prediction	RI Value	Probability
rs7412	R176C	Disease	4	0.716
rs769452	L46P	Disease	1	0.525
rs769455	R163C	Disease	7	0.865
rs28931578	R152Q	Neutral	9	0.065

4.4 Protein Stability

Stability of protein changes with a mutation in the sequence of their amino acids.The stability of proteins having a damaging effect usually decreases. To check the effect of the mutation on the stability of protein, I mutant was used. The following data is collected from this tool. The stability of selected SNPs was observed to be decreased through this tool.DDG value was also determined by using this tool. Stability status and DDG value for selected SNPs were presented in table 4.5.

In both these tables, WT is the amino acid in wild type proteins NEW is the amino acid in mutated protein, RI is the reliability index, pH is the negative logarithm of hydrogen ions concentration and T is the temperature in celsius. DDG is the difference of the G value of mutated and wild-type protein. DDG with the value of less than 0 is an indication of a decrease in stability.

TABLE 4.5: Stability status of damaging SNPs

Rs Id	Position	WT	NEW	Stability	RI	DDG	pH	T
rs7412	24	V	T	Decrease	9	-0.15	7.0	25
rs769452	18	A	C	Decrease	1	-0.28	7.0	25
rs28931577	25	E	A	Decrease	8	-0.67	7.0	25
rs28931578	26	T	A	Decrease	3	-0.13	7.0	25
rs769455	27	E	T	Decrease	7	0.11	7.0	25

4.5 Evolutionary Conservation

Some mutations are produced suddenly and are not passed on to the next generation while some mutations have evolutionary significance. They remain conserved and inherited from generation to generation. To present the evolutionary relationship of any protein, the evolutionary tree is the best one among all methods. An evolutionary Tree or phylogenetic tree is a diagram that shows evolutionary links among various organisms.

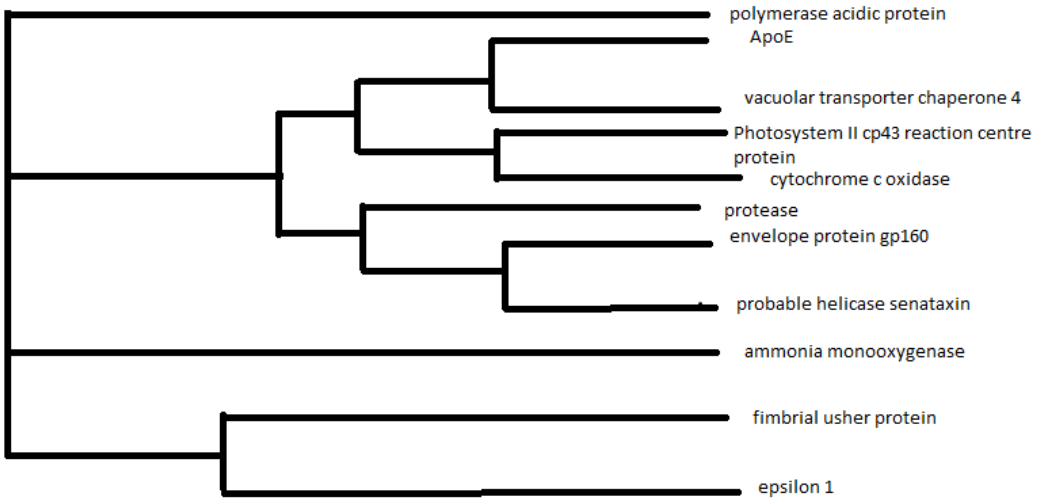

FIGURE 4.1: Evolutionary Tree of ApoE Gene

Construction of phylogenetic tree is important to arrange data of biological diversity and it also depicts the evolution of genes and proteins. Evolutionary links of Apo E were detected by using conserf server. Input was basically in the form of an amino acid sequence of the protein.

Tool presented the evolutionary relationship of protein with other proteins and multiple sequence alignment was also done by using this tool.

Multiple Sequence Alignment is a process in which two or more biomolecules with sequence of same length are compared for similarity. This sequence may be of nucleic acids or proteins. Results are represented in figure 4.1 and 4.2.

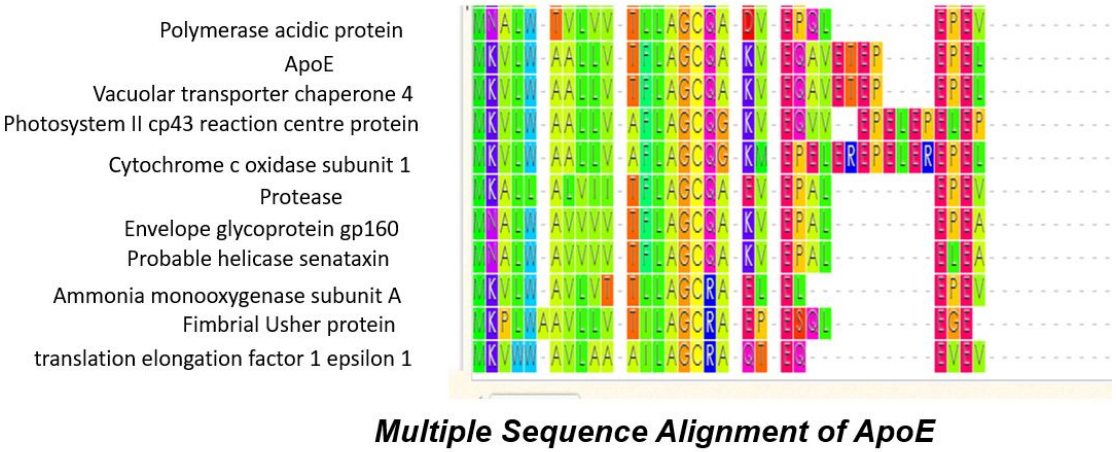

FIGURE 4.2: Multiple Sequence Alignment of ApoE

4.6 Structure prediction and validation

Any mutation in the protein may affect the structure of the protein. To determine the effect of the mutation on the structure of Apo E , the structure was built by using the phyre 2 tool. Phyre 2 tool gives information about structure, function, and mutation in protein: This tool represents a 3 D model of any protein of a given sequence. The confidence value of different SNPs was predicted by the phyre 2 tool. I-tasser is also a useful tool for predicting the 3D structure of proteins. It was initially designed to predict the structure of the protein but now its function has been extended regarding the functional analysis of proteins based on their structure. After crosschecking by I tasser following models were obtained given in figures from 4.3 to 4.7. The models proposed by I-tasser were confirmed by the save v5 server.

Figure 4.3 is the 3D model of protein with rs id 7412. The confidence value of this protein predicted by phyre2 tool is 99.3. After its confirmation by saving v5 server,the Overall quality factor value A for the above-predicted model was 93.2039.

Below **Figure 4.4** is the proposed 3D model of protein for rs id 769452 predicted by phyre2 and crosschecked by I tasser.

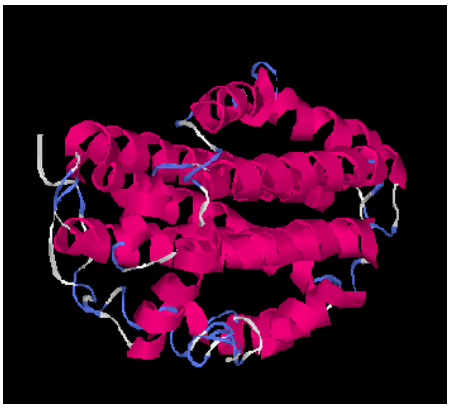

FIGURE 4.3: 3D structure of rs7412 SNP id of ApoE

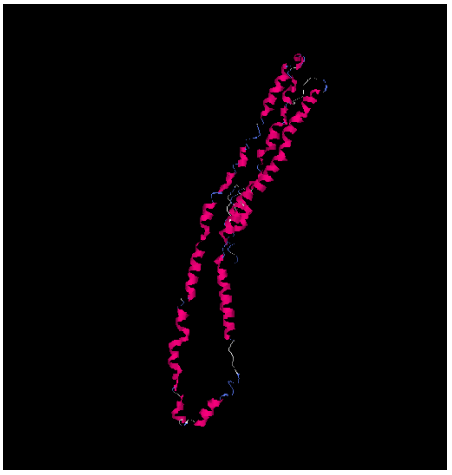

FIGURE 4.4: 3D Structure of rs769452 SNP id of ApoE

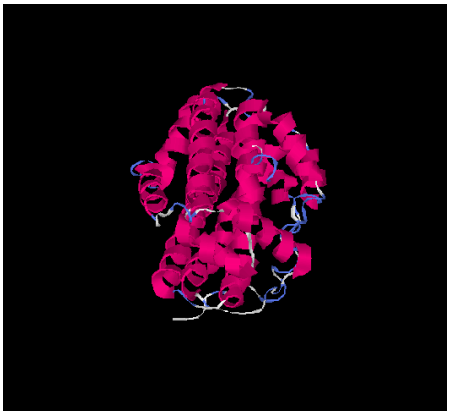

FIGURE 4.5: 3D structure of rs28931577 SNP id of ApoE

In **Figure 4.5** structure has confidence value of 99.3. This predicted structure has overall quality factor value A of 90.6149.

Figure 4.6 is the 3D model of rs769455 SNP id of ApoE. Its significance value is 99.0. Overall quality factor value A for this mutated protein structure is 91.2621.

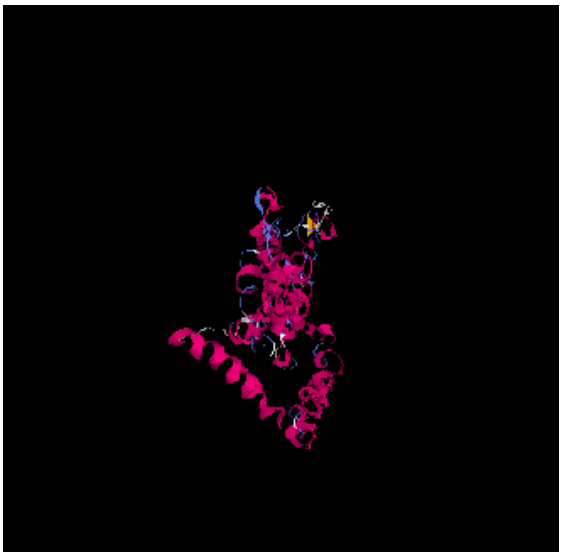

FIGURE 4.6: 3D Structure of rs769455 SNP id of ApoE

FIGURE 4.7: 3D Structure of rs28931578 SNP id of Apo E

Above **Figure 4.7** is the 3D model proposed by phyre 2 and confirmed by I tasser for protein of SNP variant with rs id 28931578. Confidence value has been calculated to be 99.3. quality factor value A of this structure is 80.7818. The differences between the observed and expected amino acid frequency were also determined by using save v5 and represented in the form of a graph in this tool. Graph showing the frequency of amino acids are represented in figures 4.8.

Figure 4.8 shows the difference in the observed and expected frequency of different amino acid in mutated proteins of ApoE.

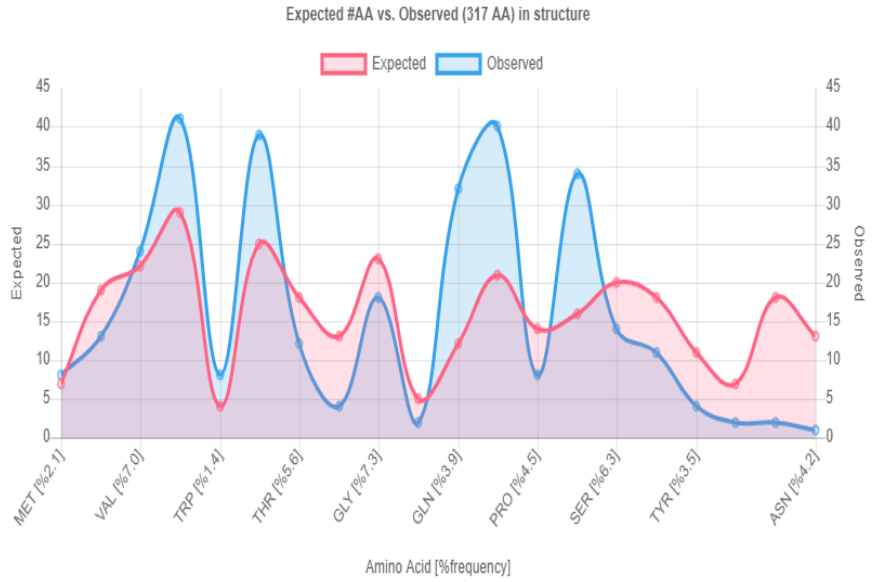

FIGURE 4.8: Graph for observed and expected amino acid frequency in wild and mutated ApoE

The difference in frequency values for different amino acids can be represented in a table 4.6 as follows.

TABLE 4.6: Difference in the observed and expected frequency of different amino acids in wild and mutated proteins

S. No	**Name of Amino acid**	**Difference in the observed and expected frequency**
1	MET	2.1%
2	VAL	7.0%
3	TRP	1.4%
4	THR	5.6%
5	GLY	7.3%
6	GLN	3.9%
7	PRO	4.5%
8	SER	6.3%
9	TYR	3.5%
10	ASN	4.2%

4.7 Structural Alignment

As the predicted structures of proteins are increasing day by day, the significance of the alignment of proteins structure has been increased. For structural alignment, the structure of wild type protein was collected from Protein Data Bank and then was given as input for structural alignment in TM align. Protein Data Bank contains the data for three-dimensional structures of proteins.

TM score for every model was below 1 which is an indication of variation in the structure of the protein. In the results, red color indicates mutated protein model and blue color indicates the wild type model. Results of TM alignment for selected SNPs of ApoE were presented in **figures 4.9 to 4.13.**

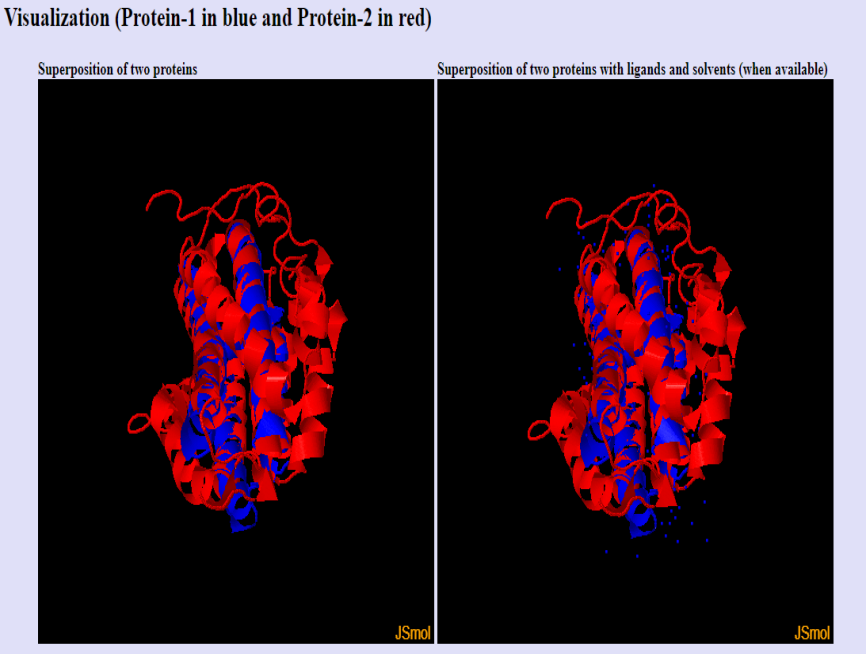

FIGURE 4.9: Result of TM alignment for model 1 of ApoE

The chain of mutated prorein proposed in this model contains 299 residues. TM score for this structure is 0.78633. Seq Id is the ration of similar amino acids to aligned amino acid which is 0.971 for this model. RMSD is the root mean square distance whose value is 2.65 for this model. Aligned region for this model is 138 amino acids long.

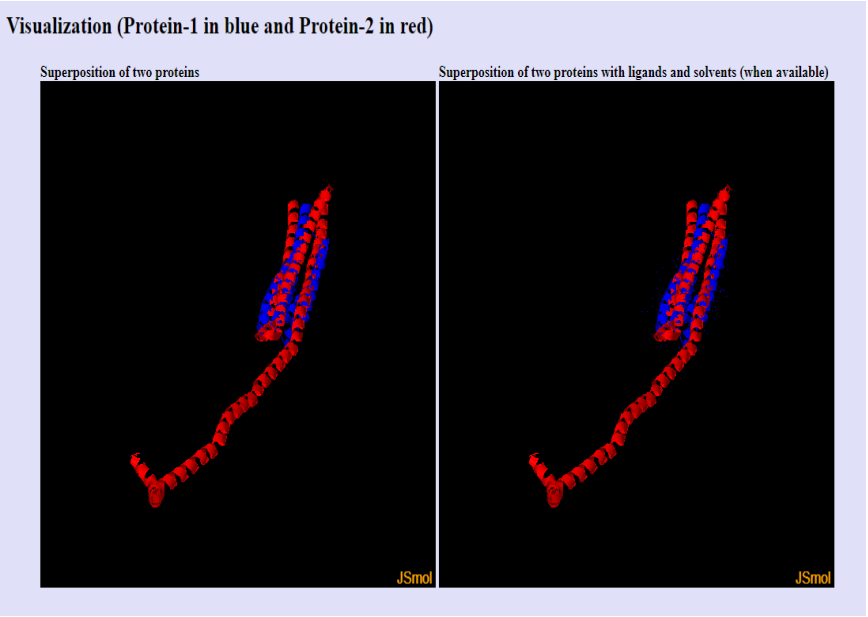

FIGURE 4.10: Result of TM alignment for model 2 of ApoE

Figure 4.10 shows the result of alignment for model 2 with wild type. Input model in this alignment contains 238 residues. 112 amino acids were aligned in both structures of the protein. RMSD value for this model is 3.33 and seq. Id is equal to 0.188.

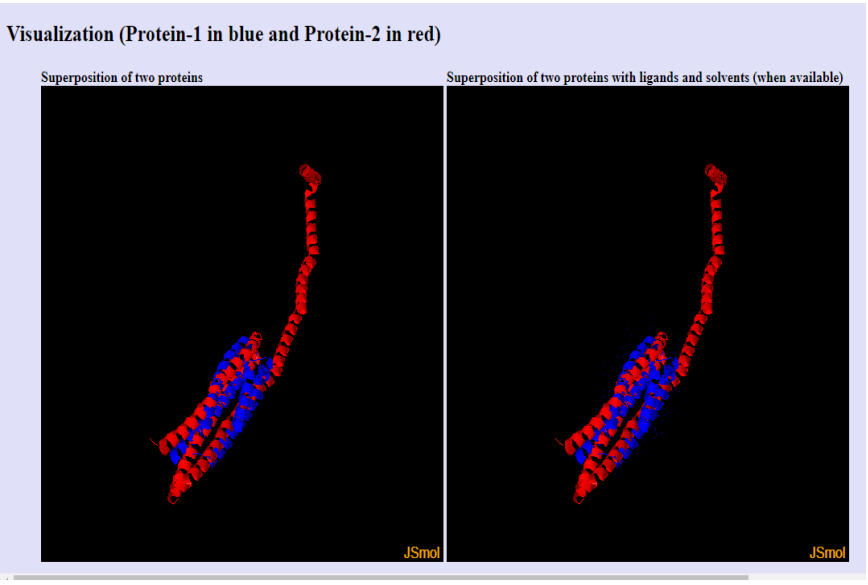

FIGURE 4.11: Result of TM alignment for model 3 of ApoE

Figure 4.11 depicts the result of TM alignment for model 3 proposed by I tasser. The aligned segment of both protein is 112 aminoacids long. This protein contains 239 residues, RMSD value for model 3 is 3.45 and seq. id is 0.188.

Figure 4.12 shows the result of alignment of model 4 with wild type protein. The input model for alignment contains 296 aminoacids. Aligned segement is 75 amino acids long. RMSD value is 3.97 and sequence id is 0.173.

Below **Figure 4.13** is the result of TM align by alignment of wild type protein with model 5. Length of protein according to this model is 183 residues and it is aligned along 136 residues. RMSD value is 2.60 and seq id is 0.721.

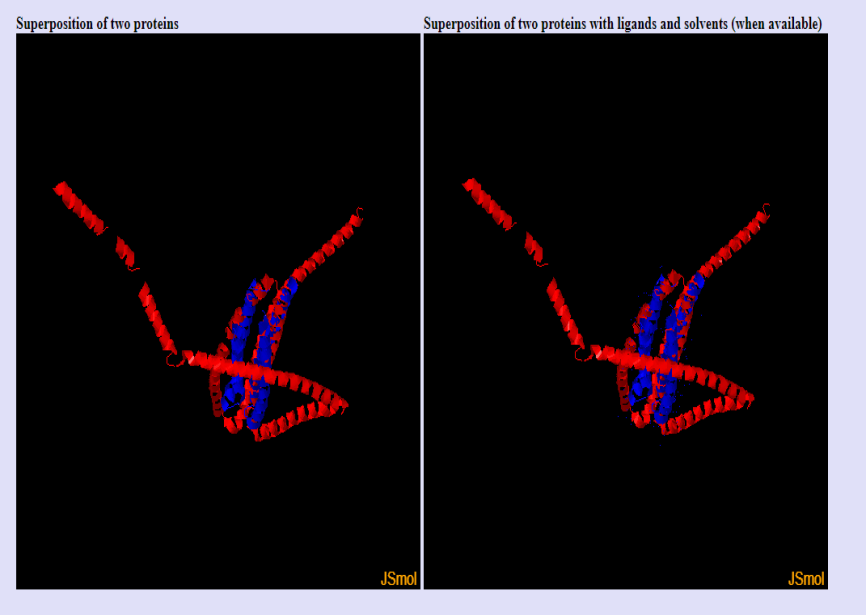

FIGURE 4.12: Result of TM alignment for model 4 of ApoE

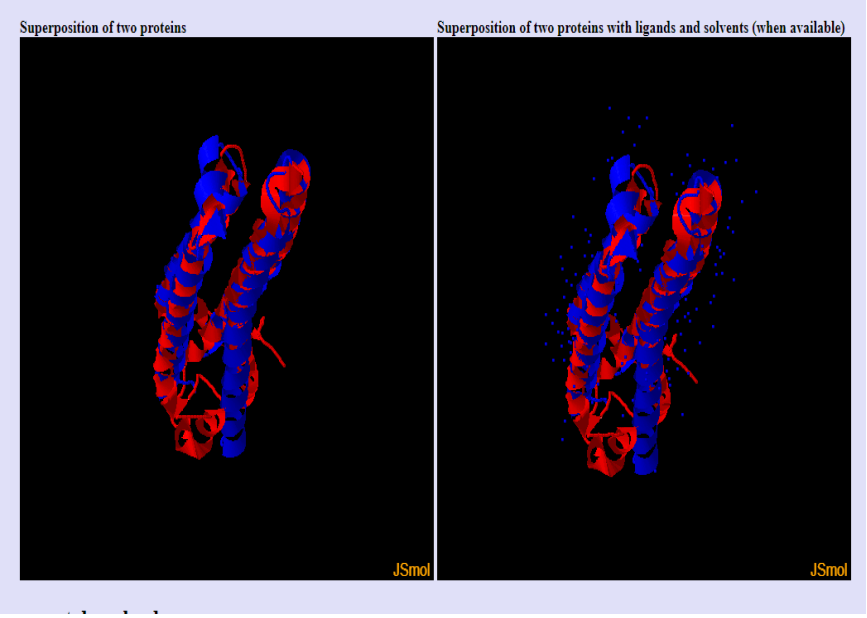

FIGURE 4.13: Result of TM alignment for model 5 of ApoE

4.8 Post-Translational Modification

Many post-translational modifications are important for the stability and functional activity of proteins. Methylatíon is also an important post-translational modification. The methylated site in any protein help to identify the problem caused by it and also the development of drugs for any disease. Methylated sites in mutated proteins were detected by pssme and then crosschecked by bpb/ppms. Methylation occur only on lysine (K) and arginine (R) residues.

The position of methylated site for both residues was detected by these tools. Methylation of protein has an important role in determining the functional activity of proteins. Therefore, information about methylation can help detect the normal or mutated functioning of proteins. SVM is an algorithm used in the study of protein interactions.

The type of Methylation also affects many characteristics of the protein. The type was detected by using a web server. The post-translational modification also involves phosphorylation other than methylation. Information about phosphorylation in mutated proteins was collected by using GPS5.0

Table 4.7: Site of methylation in mutated proteins of ApoE

A) For R residues

Protein name	**Position of site**	**Flanking residues**	**SVM Probability**
P02649	50	GQRWELALG-R -FWDYLRWVQ	0.5
P02649	190	QAGAREGAE-R -GLSAIRERL	0.84468
P02649	196	GAERGLSAI-R -ERLGPLVEQ	0.60569
P02649	209	GPLVEQGRV-R -AATVGSLAG	0.82963

Continued Table 4.7: Site of methylation in mutated proteins of ApoE

B) For K residues

Protein name	**Position \ of site**	**Flanking residues**	**SVM Probability**
P02649	2	OOOOOOOOM-K -VLWAALLVT	0.72413
P02649	87	LRALMDETM-K -ELKAYKSEL	0.64725
P02649	90	LMDETMKEL -K-AYKSELEEQ	0.74998
P02649	93	ETMKELKAY-K -SELEEQLTP	0.81515
P02649	161	RVRLASHLR-K -LRKRLLRDA	0.51941
P02649	251	RTRDRLDEV-K -EQVAEVRAK	0.81882
P02649	260	KEQVAEVRA-K -LEEQAQQIR	0.75006
P02649	280	QAEAFQARL-K -SWFEPLVED	0.87851

4.9 Interactors Identification

In our body,in the functional pathway of any protein, many other proteins also interact with that protein and affect its working. Those proteins are called interactors. In the study of the pathway of any protein, the study of interactors is very important, which not only help to identify the interactors of normal protein but also helps in the identification of affected protein in case of any disease.To identify the interactors of the apo E gene, funcoup server was used.

4.10 Construction of ApoE interaction network

Genemania is a server for network analysis. It predicts the related genes involved with one gene and also describes the pathways in a system involving those genes. Interaction Data from this server was retrieved showing source and target genes. By using source and target gene information,a directed network was generated in Gephi tool (Fig. 4.14) and statistical calculations were performed (Table 4.8).

It was observed that many genes interact with apoE but ApoE does not interact with any gene. The genes that interact with ApoE mostly involve genes related to lipid metabolism and some genes related to neurological pathways such as CNTF gene. These gene interactors confirm the involvement of ApoE in cardiovascular as well as neurological diseases.

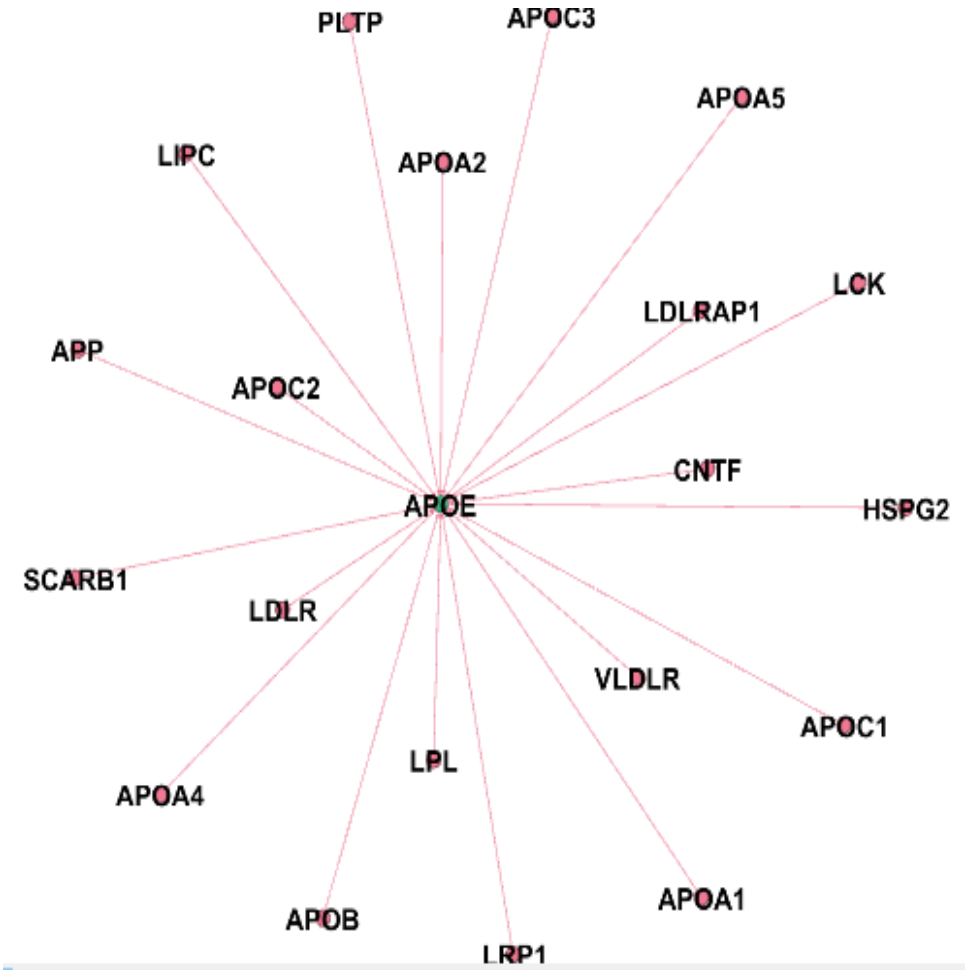

FIGURE 4.14: Network Pathway of ApoE gene predicted by Gephi

TABLE 4.8: Values of some important parameters predicted by Gephi server

Character	Value
Average Degree	0.952
Average Weighted Degree	0.476
Network Diameter	1
Graph Density	0.048
Modularity	0
Connected Components	1

4.11 Enrichment Analysis

Proteins determined to be involved with apoE were subjected to be enrichment tool for enrichment analysis. This analysis gave information about the involvement of ApoE in different pathways. This analysis uses KEGG for the prediction of pathways and NCI and reactome for crosschecking of pathways involving apoE-protein. As a result, pathways involving ApoE were determined. These tools also provided result of Fisher test. The genes having interaction with ApoE were also derected and represented in **Table 4.9 and 4.10**.

TABLE 4.9: Pathways Involving ApoE Protein Detected After Enrichment Analysis

S. No	Annotation (Pathway/ Process)	Significance of network distance distribution (XD-Score)	Significance of Overlap (Fisher test, q-value)
1	CHYLOMICRON MEDIATED LIPID TRANSPORT	6.8467	2.60E-28

Table 4.9 continued from previous page

S. No	Annotation (Pathway/ Process)	Significance of network distance distribution (XD-Score)	Significance of Overlap (Fisher test, q-value)
2	LIPOPROTEIN METABOLISM	4.631	1.10E-27
3	HDL MEDIATED LIPID TRANSPORT	1.6007	7.90E-02
4	PECAM1 INTERACTIONS	0.8644	1.00E+00
5	CD28 DEPENDENT VAV1 PATHWAY	0.7826	1.00E+00
6	REGULATION OF LIPID METABOLISM BY PEROXISOME PROLIFERATOR ACTIVATED \ RECEPTOR ALPHA	0.7021	4.20E-04
7	METABOLISM OF LIPIDS AND LIPOPROTEINS	0.5644	2.30E-15
8	CD28 DEPENDENT P13K AKT SIGNALING	0.4381	1.00E+00
9	CTLA4 INHIBITORY SIGNALING	0.4144	1.00E+00

Table 4.9 continued from previous page

S. No	Annotation (Pathway/ Process)	Significance of network distance distribution (XD-Score)	Significance of Overlap (Fisher test, q-value)
10	TRANSLOCATION OF ZAP70 TO IMMUNOLOGICAL SYNAPSE	0.4144	1.00E+00
11	FURTHER PLATELET RELEASATE	0.393	1.00E+00
12	NEF MEDIATES DOWN MODULATION OF CELL SURFACE RECEPTORS BY RECRUITING THEM TO CLATHRING ADAPT	0.393	1.00E+00
13	PHOSPHORYLATION OF CD3 AND TCR ZETA CHAINS	0.3735	1.00E+00
14	PD1 SIGNALING	0.3244	1.00E+00
15	CD28 CO STIMULATION	0.2858	1.00E+00
16	THE ROLE OF NEF IN HIV1 \REPLICATION AND DISEASE PATHOGENESIS	0.2858	1.00E+00

Table 4.9 continued from previous page

S. No	Annotation (Pathway/ Process)	Significance of network distance distribution (XD-Score)	Significance of Overlap (Fisher test, q-value)
17	GENERATION OF SECOND MESSENGER MOLECULE	0.2371	1.00E+00
18	PLATELET DEGRANULATION	0.1839	1.00E+00
19	DOWNSTREAM TCR SIGNALING	0.1737	1.00E+00
20	CELL SURFACE INTERACTIONS AT THE VASCULAR WALL	0.1622	1.00E+00
21	TCR SIGNALING	0.1144	1.00E+00
22	COSTIMULATION BY THE CD28 FAMILY	0.1028	1.00E+00
23	HEMOSTASIS	0.0997	1.00E+00
24	PLATELET ACTIVATION	0.0769	1.00E+00
25	FORMATION OF PLATELET PLUG	0.0649	1.00E+00
26	HOST INTERACTIONS OF HIV FACTORS	0.0407	1.00E+00

Table 4.9 continued from previous page

S. No	Annotation (Pathway/ Process)	Significance of network distance distribution (XD-Score)	Significance of Overlap (Fisher test, q-value)
27	SIGNALING IN IMMUNE SYSTEM	0.0203	1.00E+00
28	HIV INFECTION	0.0144	1.00E+00
29	TRANSMEMBRANE TRANSPORT OF SMALL MOLECULES	0.0066	1.00E+00
30	PPAR signaling pathway	0.7682	5.8e-06

Table 4.9 continued from same page

S. No	DataSet Size (Uploaded gene set)	Dataset Size (pathway gene set)	Dataset Size (Overlap)
1	21	17	Gene Symbol APOB, LPL APOA5, HSPG2 APOC2, APOA2 & 3 APOA4, APOA1 APOE, LIPC LDLR, LDLRAP1

Table 4.9 continued from previous page

S. No	DataSet Size (Uploaded gene set)	Dataset Size (pathway gene set)	Dataset Size (Overlap)
2	21	27	GENE SYMBOL LPL APOB HSPG2 APOA2 SCARB1 APOA4 LIPC APOA1 LDLR APOA5 APOC2 APOC3 APOE LDLRAP1
3	21	11	GENE SYMBOL SCARB1 APOA1
4	21	10	GENE SYMBOL LCK
5	21	11	GENE SYMBOL LCK

Table 4.9 continued from previous page

S. No	DataSet Size (Uploaded gene set)	Dataset Size (pathway gene set)	Dataset Size (Overlap)
6	21	61	GENE SYMBOL APOA2 APOA1 APOA5 PLTP LPL
7	21	225	GENE SYMBOL APOB SCARB1 APOA2 APOA4 APOA1 APOA5 PLTP APOC2 APOC3 APOE LPLR LIPC HSPG2 LPL LDLRAP1
8	21	19	GENE SYMBOL LCK
9	21	20	GENE SYMBOL LCK

Table 4.9 continued from previous page

S. No	DataSet Size (Uploaded gene set)	Dataset Size (pathway gene set)	Dataset Size (Overlap)
10	21	20	GENE SYMBOL LCK
11	21	21	GENE SYMBOL APOA1
12	21	21	GENE SYMBOL LCK
13	21	22	GENE SYMBOL LCK
14	21	25	GENE SYMBOL LCK
15	21	28	GENE SYMBOL LCK
16	21	28	GENE SYMBOL LCK
17	21	33	GENE SYMBOL LCK
18	21	82	GENE SYMBOL APOA1 APP
19	21	43	GENE SYMBOL LCK
20	21	91	GENE SYMBOL APOB LCK
21	21	60	GENE SYMBOL LCK

Table 4.9 continued from previous page

S. No	DataSet Size (Uploaded gene set)	Dataset Size (pathway gene set)	Dataset Size (Overlap)
22	21	65	GENE SYMBOL LCK
23	21	266	GENE SYMBOL APOB APOA1 LCK
24	21	160	APP GENE SYMBOL APOA1 APP
25	21	179	GENE SYMBOL APP APOA1
26	21	118	GENE SYMBOL LCK
27	21	322	GENE SYMBOL APOB LCK
28	21	180	GENE SYMBOL LCK
29	21	213	GENE SYMBOL APOA1
30	21	69	APOA5 APOA1 APOA2 APOC3 LPL PLTP

Among all these pathways, following pathways were selected after enrichment analysis.

Table 4.10: Pathways involving ApoE protein selected after enrichment analysis

S. No	**Annotation (Pathway/ Process)**	**Significance of network distance distribution (XD-Score)**	**Significance of Overlap (Fisher test, q-value)**
1	CHYLOMICRON MEDIATED LIPID TRANSPORT	6.8467	2.60E-28
2	LIPOPROTEIN METABOLISM	4.6310	1.1E-27
3	PPAR Signaling pathway	0.7682	5.8E-06

Table 4.10 continued from Same page

S. No	**DataSet Size (Uploaded gene set)**	**Dataset Size (pathway gene set)**	**Dataset Size (Overlap)**
1	21	17	Gene Symbol APOB, LPL APOA5, HSPG2 APOC2, APOA2 & 3 APOA4, APOA1 APOE, LIPC LDLR, LDLRAP1

Table 4.10 continued from previous page

S. No	DataSet Size (Uploaded gene set)	Dataset Size (pathway gene set)	Dataset Size (Overlap)
			GENE SYMBOL
			LPL
			APOB
			HSPG2
			APOA2
			SCARB1
			APOA4
2	21	27	LIPC
			APOA1
			LDLR
			APOA5
			APOC2
			APOC3
			APOE
			LDLRAP1
			GENE SYMBOL
			APOA5
			APOA1
3	21	69	APOA2
			APOC3
			LPL
			PLTP

The first two pathways were selected in reactone database enrichment analysis and the last pathway was selected in kegg database enrichment analysis. The value of Absolute Pearson correlation between XD-scores and Fisher q-values is

0.75 for reactone and 0.83 for kegg database analysis. Whereas threshold value for XD-score significance is 3.02 for reactome and 0.59 for keg database enrichment analysis. So in future to study any disease involving these pathways can be studied by effects of ApoE in that disease.

Chapter 5

Conclusions and Recommendations

ApoE gene has a great significance in genetic research because it is considered as a risk factor for many diseases. This research was designed to detect the deleterious SNPs of ApoE by using different bioinformatics tools. The damaging effects were confirmed and structures of mutated proteins were aligned. Furthermore, interactors were also identified to detect all possible pathways involving Apo E gene and which pathways can be affected by any change in the nucleotide sequence of ApoE. In this study first objective was achieved by using the sift server and polyphen tool. SNPs for which protein sequences have been determined were given as input and their effects were confirmed. By using these tools, five SNPs have been found to have deleterious effects. GO Terms were found by GO term finder EBI and then by using SNP & GO structural and functional properties of mutated proteins were determined.Furthermore, their deleterious effects were confirmed by checking the stability of mutated proteins produced as a result of these SNPs. I mutant was used to confirm the deleterious effects of proteins and consurf server was used to detect the evolutionary conservation of proteins. Second objective was achieved by determining and validating the secondary structure of mutated proteins by phyre2, I-tasser and Save V5 tool. The stability of protein structure was also determined by using these tools. Information about the stability of protein structure further

confirms the deleterious effects of these mutated proteins. Then these structures were aligned with wild protein by using TM Align tool. This tool detected the similar and different regions in wild protein structure and models detected by earlier tools. Third objective was achieved by using funcoup server for interactors identification. Analysis was done by using gephi tool. At last , enrichment analysis was performed to determine the most related pathways involving ApoE gene. To target this gene in drugs it is necessary to identify pathways involving ApoE gene. Drugs against these diseases can be designed by targeting these pathways. In the modern era, major research is about personalized medicine. This research can be helpful in this field regarding the ApoE genotype. This study can be helpful in future work of drug development against the diseases involving variation in plasma lipid level after in-vitro validation of these result.

Bibliography

[1] R. W. Mahley, "Apolipoprotein E: Cholesterol Transport," Science (80-)., vol. 240, pp. 622–630, 1988.

[2] Q. Xu, A. Bernardo, D. Walker, T. Kanegawa, R. W. Mahley, and Y. Huang, "Profile and regulation of apolipoprotein E (ApoE) expression in the CNS in mice with targeting of green fluorescent protein gene to the ApoE locus," J. Neurosci., vol. 26, no. 19, pp. 4985–4994, 2006, doi: 10.1523/JNEUROSCI.5476-05.2006.

[3] K. Ueno et al., "Supplementary Information of 'Discovery of superconductivity in $KTaO_3$ by electrostatic carrier doping,'" Nat. Nanotechnol., vol. 6, no. 7, pp. 408–12, 2011, doi: 10.1016/j.nbd.2014.08.025.Apolipoprotein.

[4] M. Xhignesse, S. Lussier-Cacan, C. F. Sing, A. M. Kessling, and J. Davignon, "Influences of common variants of apolipoprotein E on measures of lipid metabolism in a sample selected for health," Arterioscler. Thromb. Vasc. Biol., vol. 11, no. 4, pp. 1100–1110, 1991, doi: 10.1161/01.ATV.11.4.1100.

[5] P. Abondio et al., "The genetic variability of APOE in different human populations and its implications for longevity," Genes (Basel)., vol. 10, no. 3, 2019, doi: 10.3390/genes10030222.

[6] C. H. Schilling, S. Schuster, B. O. Palsson, and R. Heinrich, "Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era," Biotechnol. Prog., vol. 15, no. 3, pp. 296–303, 1999, doi: 10.1021/bp990048k.

[7] S. Schuster, D. A. Fell, and T. Dandekar, "Nature Biotechnology 2000 Schuster," vol. 18, no. March, pp. 1–7, 2000, [Online]. Available: papers2://publication/uuid/00A659FC-63C2-438F-880F-F7D7070A1E4B.

[8] S. M. Lingala and M. G. M. Mhs. Ghany, "The Biophysics and Cell Biology of Lipid Droplets HHS Public Access," vol. 25, no. 3, pp. 289–313, 2016, doi: 10.1038/nrm3699.The.

[9] T. Fujimoto and R. G. Parton, "Not just fat: The structure and function of the lipid droplet," Cold Spring Harb. Perspect. Biol., vol. 3, no. 3, pp. 1–17, 2011, doi: 10.1101/cshperspect.a004838.

[10] K. Papsdorf and A. Brunet, "Linking Lipid Metabolism to Chromatin Regulation in Aging," Trends Cell Biol., vol. 29, no. 2, pp. 97–116, 2019, doi: 10.1016/j.tcb.2018.09.004.

[11] R. M. Corbo and R. Scacchp, "Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele?," Ann. Hum. Genet., vol. 63, no. 4, pp. 301–310, 1999, doi: 10.1017/S0003480099007575.

[12] J. H. Stengård et al., "Contributions of 18 additional DNA sequence variations in the gene encoding apolipoprotein E to explaining variation in quantitative measures of lipid metabolism," Am. J. Hum. Genet., vol. 71, no. 3, pp. 501–517, 2002, doi: 10.1086/342217.

[13] C. Ehnholm, M. Lukka, T. Kuusi, E. Nikkilä, and G. Utermann, "Apolipoprotein E polymorphism in the Finnish population: Gene frequencies and relation to lipoprotein concentrations," J. Lipid Res., vol. 27, no. 3, pp. 227–235, 1986, doi: 10.1016/s0022-2275(20)38833-7.

[14] H. H. Bock, Y. Jossin, P. May, O. Bergner, and J. Herz, "Apolipoprotein E receptors are required for reelin-induced proteasomal degradation of the neuronal adaptor protein disabled-1," J. Biol. Chem., vol. 279, no. 32, pp. 33471–33479, 2004, doi: 10.1074/jbc.M401770200.

[15] R. Couderc, F. Mahieux, S. Bailleul, G. Fenelon, R. Mary, and J. Fermanian, "Prevalence of apolipoprotein E phenotypes in ischemic cerebrovascular disease: A case-control study," Stroke, vol. 24, no. 5, pp. 661–664, 1993, doi: 10.1161/01.STR.24.5.661.

[16] A. Ramadan, B. M. Foda, A. S. Noha, O. Refaat, A. A. Saleh, and A. Fawzy, "Association analysis of ApoE gene polymorphisms among Egyptian patients with Alzheimer's disease," Gene Reports, vol. 14, no. October, pp. 65–71, 2019, doi: 10.1016/j.genrep.2018.10.017.

[17] L. A. Gauri et al., "Study of apoe gene and d2s439 marker in patients of rheumatoid arthritis and their correlation with severity of disease: A case control study," J. Assoc. Physicians India, vol. 67, no. July, pp. 43–48, 2019.

[18] N. S. Vučinić, J. Drljača, K. Bačulov, and K. D. Jeremić, "The E4 allele of the apoE gene acts as a determinant of Metabolic syndrome in men," vol. 41, no. 1, pp. 45–50, 2019, doi: 10.5281/zenodo.3250178.

[19] S. Cetinkalp et al., "Apolipoprotein e gene polymorphism and polycystic ovary syndrome patients in Western Anatolia, Turkey," J. Assist. Reprod. Genet., vol. 26, no. 1, pp. 1–6, 2009, doi: 10.1007/s10815-008-9280-8.

[20] M. Cui et al., "Risk Assessment of Sarcopenia in Patients With Type 2 Diabetes Mellitus Using Data Mining Methods," Front. Endocrinol. (Lausanne)., vol. 11, 2020, doi: 10.3389/fendo.2020.00123.

[21] D. Iacono and G. C. Feltis, "Impact of Apolipoprotein E gene polymorphism during normal and pathological conditions of the brain across the lifespan," Aging (Albany. NY)., vol. 11, no. 2, pp. 787–816, 2019, doi: 10.18632/aging.101757.

[22] A. TSARBOPOULOS, M. GROSS, and I. JARDINE, "Rapid identification of calbindin-D," Biomed. Environ. Mass Spectrom., vol. 18, no. 6, pp. 387–393, 1989.

[23] M. Dzietko et al., "Apolipoprotein E gene polymorphisms and intraventricular haemorrhage in infants born preterm: a large prospective multicentre cohort study," Dev. Med. Child Neurol., vol. 61, no. 3, pp. 337–342, 2019, doi: 10.1111/dmcn.13987.

[24] L.-M. Dong and K. H. Weisgraber, "Human Apolipoprotein E4 Domain Interaction," J. Biol. Chem., vol. 271, no. 32, pp. 19053–19057, 1996, doi: 10.1074/jbc.271.32.19053.

[25] K. H. Weisgraber, R. W. Mahley, R. C. Kowalnii, J. Herzli, J. L. Goldsteinq, and M. S. Brownn, "Apolipoprotein C-I Modulates the Interaction of Apolipoprotein E with @-Migrating Very Low Density Lipoproteins (P-VLDL) and Inhibits Binding of &VLDL to Low Density Lipoprotein Receptor-related Protein*," 1990.

[26] L. M. Dong et al., "Human apolipoprotein E. Role of arginine 61 in mediating the lipoprotein preferences of the E3 and E4 isoforms," J. Biol. Chem., vol. 269, no. 35, pp. 22358–22365, 1994, doi: 10.1016/s0021-9258(17)31797-0.

[27] J. R. Wetterau, L. P. Aggerbeck, S. C. Rall, and K. H. Weisgraber, "Human apolipoprotein E3 in aqueous solution. 1. Evidence for two structural domains," J. Biol. Chem., vol. 263, no. 13, pp. 6240–6248, 1988, doi: 10.1016/s0021-9258(18)68778-2.

[28] K. H. Weisgraber, "Apolipoprotein E distribution among human plasma lipoproteins: Role of the cysteine-arginine interchange at residue 112," J. Lipid Res., vol. 31, no. 8, pp. 1503–1511, 1990, doi: 10.1016/s0022-2275(20)42621-5.

[29] R. W. Mahley, T. L. Innerarity, S. C. Rall, and K. H. Weisgraber, "Plasma lipoproteins: Apolipoprotein structure and function," J. Lipid Res., vol. 25, no. 12, pp. 1277–1294, 1984, doi: 10.1016/s0022-2275(20)34443-6.

[30] W. J. McConathy and C. S. Wang, "Inhibition of lipoprotein lipase by the receptor-binding domain of apolipoprotein E," FEBS Lett., vol. 251, no. 1–2, pp. 250–252, 1989, doi: 10.1016/0014-5793(89)81464-4.

[31] T. L. Innerarity, K. H. Weisgraber, K. S. Arnold, S. C. Rall, and R. W. Mahley, "Normalization of receptor binding of apolipoprotein E2. Evidence for modulation of the binding site conformation," J. Biol. Chem., vol. 259, no. 11, pp. 7261–7267, 1984, doi: 10.1016/s0021-9258(17)39866-6.

[32] M. R. Wardell, K. H. Weisgraber, L. M. Havekes, and S. C. Rall, "Apolipoprotein E3-Leiden contains a seven-amino acid insertion that is a tandem repeat of residues 121-127," J. Biol. Chem., vol. 264, no. 35, pp. 21205–21210, 1989, doi: 10.1016/s0021-9258(19)30067-5.

[33] R. E. Gregg and H. B. Brewer, "In vivo metabolism of apolipoprotein E in humans," Methods Enzymol., vol. 129, no. C, pp. 482–497, 1986, doi: 10.1016/0076-6879(86)29087-4.

[34] A. Steinmetz, C. Jakobs, S. Motzny, and H. Kaffarnik, "Differential distribution of apolipoprotein E isoforms in human plasma lipoproteins," Arteriosclerosis, vol. 9, no. 3, pp. 405–411, 1989, doi: 10.1161/01.atv.9.3.405.

[35] J. Davignon, R. E. Gregg, and C. F. Sing, "Review Apolipoprotein E Polymorphism and Atherosclerosis," no. February, pp. 1–21, 1988.

[36] L. Ramos, J. Ferreira, F. Almeida, L. Helena, and S. Pimassoni, "The combined risk effect among BIN1 , CLU , and APOE genes in Alzheimer ' s disease," vol. 20180320, pp. 1–7, 2020.

[37] R. W. Mahley, K. H. Weisgraber, and Y. Huang, "Apolipoprotein E: Structure determines function, from atherosclerosis to Alzheimer's disease to AIDS," J. Lipid Res., vol. 50, no. SUPPL., pp. 183–188, 2009, doi: 10.1194/jlr.R800069-JLR200.

[38] D. J. Shaw, J. D. Brook, A. L. Meredith, H. G. Harley, M. Sarfarazi, and P. S. Harper, "Gene mapping and chromosome 19," J. Med. Genet., vol. 23, no. 1, pp. 2–10, 1986, doi: 10.1136/jmg.23.1.2.

[39] E. S. Fenjves, D. A. Gordon, L. K. Pershingt, D. L. Williams, and L. B. Taichman, "Systemic distribution of apolipoprotein E secreted by grafts of epidermal keratinocytes: Implications for epidermal function and gene therapy," vol. 86, no. November, pp. 8803–8807, 1989.

[40] J. W. Mclean, N. A. Elshourbagy, D. J. Chang, R. W. Mahley, and J. M. Taylor, "Human Apolipoprotein E mRNA," no. 21, pp. 6498–6504, 1984, doi: 10.1016/S0021-9258(20)82169-3.

[41] J. Simon-Sanchez et al., "Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals," Hum. Mol. Genet., vol. 16, no. 1, pp. 1–14, 2007, doi: 10.1093/hmg/ddl436.

[42] S. Teng, E. Michonova-Alexova, and E. Alexov, "Approaches and Resources for Prediction of the Effects of Non-Synonymous Single Nucleotide Polymorphism on Protein Function and Interactions," Curr. Pharm. Biotechnol., vol. 9, no. 2, pp. 123–133, 2008, doi: 10.2174/138920108783955164.

[43] S. Mooney, "Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis," Brief. Bioinform., vol. 6, no. 1, pp. 44–56, 2005, doi: 10.1093/bib/6.1.44.

[44] M. Cargill et al., "Characterization of single-nucleotide polymorphisms in coding regions of human genes," Nat. Genet., vol. 22, no. 3, pp. 231–238, 1999, doi: 10.1038/10290.

[45] N. A. Elshourbagy, J. W. McLean, R. W. Mahley, and J. M. Taylor, "Apolipoprotein E mRNA is relatively abundant in the brain, as well as in the liver and other tissues of marmosets and rats," Fed. Proc., vol. 43, no. 6, pp. 203–207, 1984.

[46] R. W. Mahley, K. H. Weisgraber, and Y. Huang, "Apolipoprotein E: Structure determines function, from atherosclerosis to Alzheimer's disease to AIDS," J. Lipid Res., vol. 50, no. SUPPL., pp. S183–S188, 2009, doi: 10.1194/jlr.R800069-JLR200.

[47] W. Xiong, E. Zsigmond, A. M. Gotto, K. Y. Lei, and L. Chan, "Locating a low density lipoprotein-targeting domain of human apolipoprotein B-100 by expressing a minigene construct in transgenic mice," J. Biol. Chem., vol. 266, no. 31, pp. 20893–20898, 1991, doi: 10.1016/s0021-9258(18)54794-3.

[48] G. S. Getz and C. A. Reardon, "Apoprotein E as a lipid transport and signaling protein in the blood, liver, and artery wall," J. Lipid Res., vol. 50, no. SUPPL., pp. 156–161, 2009, doi: 10.1194/jlr.R800058-JLR200.

[49] Y. E. Vanderbeeken, J. Duchateau, H. Collet, M. Gregoire, B. Vandermeersh, and a. Lucas, "Receptor Expression," Immunopharmacol. Immunotoxicol., vol. 14, no. 3, pp. 491–513, 1992.

[50] R. Pechlaner et al., "Very-Low-Density Lipoprotein–Associated Apolipoproteins Predict Cardiovascular Events and Are Lowered by Inhibition of APOC-III," J. Am. Coll. Cardiol., vol. 69, no. 7, pp. 789–800, 2017, doi: 10.1016/j.jacc.2016.11.065.

[51] S. Bellosta et al., "Macrophage-specific expression of human apolipoprotein E reduces atherosclerosis in hypercholesterolemic apolipoprotein E-null mice," J. Clin. Invest., vol. 96, no. 5, pp. 2170–2179, 1995, doi: 10.1172/JCI118271.

[52] K. Ali, M. Middleton, E. Puré, and D. J. Rader, "Apolipoprotein E suppresses the type I inflammatory response in vivo," Circ. Res., vol. 97, no. 9, pp. 922–927, 2005, doi: 10.1161/01.RES.0000187467.67684.43.

[53] P. Van Den Elzen et al., "Apolipoprotein-mediated pathways of lipid antigen presentation," Nature, vol. 437, no. 7060, pp. 906–910, 2005, doi: 10.1038/nature04001.

[54] D. Kothapalli et al., "Antimitogenic effects of HDL and APOE mediated by Cox-2-dependent IP activation," J. Clin. Invest., vol. 113, no. 4, pp. 609–618, 2004, doi: 10.1172/JCI19097.

[55] Z. H. Huang, C. A. Reardon, G. S. Getz, N. Maeda, and T. Mazzone, "Selective suppression of adipose tissue apoE expression impacts systemic metabolic phenotype and adipose tissue inflammation," J. Lipid Res., vol. 56, no. 2, pp. 215–226, 2015, doi: 10.1194/jlr.M050567.

[56] J. M. Arbones-Mainar, L. A. Johnson, M. K. Altenburg, and N. Maeda, "Differential modulation of diet-induced obesity and adipocyte functionality by human apolipoprotein E3 and E4 in mice," Int. J. Obes., vol. 32, no. 10, pp. 1595–1605, 2008, doi: 10.1038/ijo.2008.143.

[57] F. W. Robertson and A. M. Cumming, "Effects of apoprotein E polymorphism on serum lipoprotein concentration," Arteriosclerosis, vol. 5, no. 3, pp. 283–292, 1985, doi: 10.1161/01.atv.5.3.283.

[58] C. F. Sing and J. Davignon, "Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation," Am. J. Hum. Genet., vol. 37, no. 2, pp. 268–285, 1985.

[59] T. Kuusi et al., "Apoprotein E polymorphism and coronary artery disease. Increased prevalence of apolipoprotein E-4 in angiographically verified coronary patients," Arteriosclerosis, vol. 9, no. 2, pp. 237–241, 1989, doi: 10.1161/01.atv.9.2.237.

[60] I. Type and A. Abdulrahman, "Influence of copper deficiency on plasma lipoproteins and the development of enlarged plasma volume and cholesterol pool size," 2021.

[61] E. H. Corder et al., "Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families," Science (80-.)., vol. 261, no. 5123, pp. 921–923, 1993, doi: 10.1126/science.8346443.

[62] W. J. Strittmatter et al., "Apolipoprotein E: High-avidity binding to , B-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease 4℃ ," vol. 90, no. March, pp. 1977–1981, 1993.

[63] M. W. Gonzalez and M. G. Kann, "Chapter 4: Protein Interactions and Disease," PLoS Comput. Biol., vol. 8, no. 12, 2012, doi: 10.1371/journal.pcbi.1002819.

[64] S. Jaeger and P. Aloy, "From protein interaction networks to novel therapeutic strategies," IUBMB Life, vol. 64, no. 6, pp. 529–537, 2012, doi: 10.1002/iub.1040.

[65] A. L. Barabási and Z. N. Oltvai, "Network biology: Understanding the cell's functional organization," Nat. Rev. Genet., vol. 5, no. 2, pp. 101–113, 2004, doi: 10.1038/nrg1272.

[66] P. Grindrod and M. Kibble, "Review of uses of network and graph theory concepts within proteomics," Expert Rev. Proteomics, vol. 1, no. 2, pp. 229–238, 2004, doi: 10.1586/14789450.1.2.229.

[67] H. Yu et al., "NIH Public Access," vol. 322, no. 5898, pp. 104–110, 2009, doi: 10.1126/science.1158684.High.

[68] S. H. Yook, Z. N. Oltvai, and A. L. Barabási, "Functional and topological characterization of protein interaction networks," Proteomics, vol. 4, no. 4, pp. 928–942, 2004, doi: 10.1002/pmic.200300636.

[69] M. G. Kann, "Protein interactions and disease: Computational approaches to uncover the etiology of diseases," Brief. Bioinform., vol. 8, no. 5, pp. 333–346, 2007, doi: 10.1093/bib/bbm031.

[70] S. Erten et al., "Identifying stage-specific protein subnetworks for colorectal cancer," BMC Proc., vol. 6, no. Suppl 7, pp. 1–12, 2012, doi: 10.1186/1753-6561-6-S7-S1.

[71] R. Singh, J. Xu, and B. Berger, "Struct2Net: Integrating structure into protein-protein interaction prediction," Proc. Pacific Symp. Biocomput. 2006, PSB 2006, pp. 403-414, 2006, doi: 10.1142/9789812701626_0037.

[72] S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles, "58345 151..202," vol. 420, no. November, pp. 1-4, 2002, [Online]. Available: papers3://publication/uuid/2D36D7A6-A81D-467C-BCB4-C5DCA8EE3F86.

[73] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson, "Comparison of network-based pathway analysis methods," Trends Biotechnol., vol. 22, no. 8, pp. 400–405, 2004, doi: 10.1016/j.tibtech.2004.06.010.

[74] S. Schuster, T. Dandekar, and D. A. Fell, "Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering," Trends Biotechnol., vol. 17, no. 2, pp. 53–60, 1999, doi: 10.1016/S0167-7799(98)01290-6.

[75] S. Draghici et al., "A systems biology approach for pathway level analysis," Genome Res., vol. 17, no. 10, pp. 1537–1545, 2007, doi: 10.1101/gr.6202607.

[76] Q. Zhang, J. Li, H. Xie, H. Xue, and Y. Wang, "A network-based pathway-expanding approach for pathway analysis," BMC Bioinformatics, vol. 17, no. Suppl 17, 2016, doi: 10.1186/s12859-016-1333-x.

[77] S. Klamt and J. Stelling, "Combinatorial complexity of pathway analysis in metabolic networks," Mol. Biol. Rep., vol. 29, no. 1–2, pp. 233–236, 2002, doi: 10.1023/A:1020390132244.

[78] P. Kumar, S. Henikoff, and P. C. Ng, "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm," Nat. Protoc., vol. 4, no. 7, pp. 1073–1082, 2009, doi: 10.1038/nprot.2009.86.

[79] I. Adzhubei, D. M. Jordan, and S. R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2, vol. 2, no. SUPPL.76. 2013.

[80] E. Capriotti, R. Calabrese, P. Fariselli, P. L. Martelli, R. B. Altman, and R. Casadio, "WS-SNPs&GO: a web server for predicting the deleterious effect of

human protein variants using functional annotation.," BMC Genomics, vol. 14 Suppl 3, no. Suppl 3, 2013, doi: 10.1186/1471-2164-14-s3-s6.

[81] E. Capriotti and R. B. Altman, "Improving the prediction of disease-related variants using protein three-dimensional structure," BMC Bioinformatics, vol. 12, no. SUPPL. 4, 2011, doi: 10.1186/1471-2105-12-S4-S3.

[82] B. Li et al., "Automated inference of molecular mechanisms of disease from amino acid substitutions," Bioinformatics, vol. 25, no. 21, pp. 2744–2750, 2009, doi: 10.1093/bioinformatics/btp528.

[83] J. Cheng, A. Randall, and P. Baldi, "Prediction of protein stability changes for single-site mutations using support vector machines," Proteins Struct. Funct. Genet., vol. 62, no. 4, pp. 1125–1132, 2006, doi: 10.1002/prot.20810.

[84] E. Capriotti, P. Fariselli, and R. Casadio, "I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure," Nucleic Acids Res., vol. 33, no. SUPPL. 2, pp. 306–310, 2005, doi: 10.1093/nar/gki375.

[85] H. Ashkenazy et al., "ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules," Nucleic Acids Res., vol. 44, no. W1, pp. W344–W350, 2016, doi: 10.1093/nar/gkw408.

[86] G. Celniker et al., "ConSurf: Using evolutionary data to raise testable hypotheses about protein function," Isr. J. Chem., vol. 53, no. 3–4, pp. 199–206, 2013, doi: 10.1002/ijch.201200096.

[87] A. Manuscript, "Function Prediction," vol. 5, no. 4, pp. 45–67, 2005, doi: 10.1201/9781420035001.ch3.

[88] Y. Zhang, "I-TASSER server for protein 3D structure prediction," BMC Bioinformatics, vol. 9, pp. 1–8, 2008, doi: 10.1186/1471-2105-9-40.

[89] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, "The I-TASSER suite: Protein structure and function prediction," Nat. Methods, vol. 12, no. 1, pp. 7–8, 2014, doi: 10.1038/nmeth.3213.

[90] Y. Zhang and J. Skolnick, "Scoring function for automated assessment of protein structure template quality," Proteins Struct. Funct. Genet., vol. 57, no. 4, pp. 702–710, 2004, doi: 10.1002/prot.20264.

[91] Y. Zhang and J. Skolnick, "TM-align: A protein structure alignment algorithm based on the TM-score," Nucleic Acids Res., vol. 33, no. 7, pp. 2302–2309, 2005, doi: 10.1093/nar/gki524.

[92] W. Deng, Y. Wang, L. Ma, Y. Zhang, S. Ullah, and Y. Xue, "Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins," Brief. Bioinform., vol. 18, no. 4, pp. 647–658, 2017, doi: 10.1093/bib/bbw041.

[93] A. Alexeyenko and E. L. L. Sonnhammer, "Global networks of functional coupling in eukaryotes from comprehensive data integration," Genome Res., vol. 19, no. 6, pp. 1107–1116, 2009, doi: 10.1101/gr.087528.108.

[94] J. Montojo, K. Zuberi, H. Rodriguez, G. D. Bader, and Q. Morris, "GeneMANIA: Fast gene network construction and function prediction for Cytoscape," F1000Research, vol. 3, pp. 1–7, 2014, doi: 10.12688/f1000research.4572.1.

[95] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks. BT - International AAAI Conference on Weblogs and Social," Int. AAAI Conf. Weblogs Soc. Media, no. July, pp. 361–362, 2009, doi: 10.13140/2.1.1341.1520.

[96] D. Zhang et al., "AllEnricher: A comprehensive gene set function enrichment tool for both model and non-model species," BMC Bioinformatics, vol. 21, no. 1, pp. 1–6, 2020, doi: 10.1186/s12859-020-3408-y.