BioMedInformatics

MDPI

Article

Multiple Bio-Computational Tools Emerge as Valid Approach in the Assessment of Apolipoproteins Pathogenicity Related Mutations

Giorgia Francesca Saraceno and Erika Cione *

Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, 87036 Rende, Italy; giorgiafrancesca.saraceno@unical.it

* Correspondence: erika.cione@unical.it

Abstract: Background: Critical studies have unwaveringly established the importance of peculiar single-nucleotide polymorphisms (SNPs) in apolipoproteins (Apos) genes as genetic risk factors for dyslipidemias and their related comorbidities. In this study, we employed in silico approaches to analyze mutations in Apos. **Methods:** A comprehensive set of computational tools was utilized. The tools for predictions derived from sequence analysis were: SIFT, PolyPhen-2, FATHMM and SNPs&GO; The tools for structure analysis were: mCSM, DynaMut2, MAESTROweb, and PremPS; for prediction of pathogenic potential were: MutPred2, and PhD-SNP; for profiling of aggregation propensity were: Camsol, and Aggrescan3D 2.0, and lastly, for residual frustration analysis, the Frustratometer was used. These approaches assess variant effects on protein structure, stability, and function. **Results:** We identified seventeen SNPs in total, twelve for ApoB, one for ApoC2, one for ApoC3, and three for ApoE, representing 70%, 6%, 6% and 18%, respectively. The pathogenity of ApoE, was highlighted in two SNPs the rs769452 with amino acid replacement L46P, and rs769455 with amino acid replacement R163C. The aggregation/solubility analysis revealed that the L46P leads to a decrease in ApoE aggregation. The R163C, showed a decrease in solubility in one of two tools used, resulting in destabilizing effects altering its solubility. **Conclusions:** The two mutations in ApoE studied with the in silico methodologies identified clinically significant genetic variants, highlighting the robustness of the integrated approach. The future direction of the research is to create a multiplex panel with the SNPs identified here in APOE and expanding to other proteins to have a panel genetic risk assessment and disease prediction in which ApoE correlates.

Keywords: apolipoproteins; single nucleotide polymorphism; in silico

check for updates

Academic Editor: Hans Binder

Received: 23 January 2025
Revised: 4 March 2025
Accepted: 18 March 2025
Published: 20 March 2025

Citation: Saraceno, G.F.; Cione, E. Multiple Bio-Computational Tools Emerge as Valid Approach in the Assessment of Apolipoproteins Pathogenicity Related Mutations. *BioMedInformatics* **2025**, *5*, 16. https://doi.org/10.3390/biomedinformatics5010016

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Apolipoproteins (Apos) are essential proteins that associate with lipids to create lipoproteins. These complexes are essential for lipid metabolism, facilitating lipid transport through the lymphatic and circulatory systems [1]. The Apos protein family includes over twenty identified members playing also roles beyond lipid metabolism since their genetic and genomic phenotypes correlate to predispose humans to various disorders. A significant aspect of these genetic risks is attributed to Apos SNPs, which contribute to the functional diversity of Apos. These variations are closely linked to differences in disease susceptibility and progression, particularly in vascular conditions [1]. Polymorphisms in the Apo gene family are associated with various diseases. For example, ApoE genotypes are linked to circulating lipid levels, increased risk of coronary heart disease,

and cardiovascular disease (CVD) [1–3]. Individuals with the ε4 homozygous genotype are at a higher risk of developing Alzheimer's disease (AD) [4–6]. Variants in ApoA5 are associated with hypertriglyceridemia [7], while ApoA polymorphisms are connected to familial hypercholesterolemia [8]. Additionally, alterations in the ApoB gene are linked to an increased risk of ischemic heart disease [9]. In this scenario, the composition of Apos not only determines the cargo, processing, and transport efficiency of lipoproteins but also serves as a critical factor in disease risk assessment. By integrating the genetic and functional variability of Apos, this approach highlights their central role in disease heterogeneity and clinical assessment [10,11]. A particularly significant area of research focuses on single nucleotide variants that result in amino acid substitutions in proteins, known as missense mutations, which are associated with more than half of all known inherited diseases [12,13]. This study aims to investigate, using in silico tools, the impact of single nucleotide mutations on the stability of Apos.

2. Materials and Methods

2.1. Data Retrieval and Web Resources

We searched the dbSNP database for missense mutations in several Apos, including ApoA types 1, 2, 4, and 5; ApoB, ApoC types1, 2, 3 and 4; ApoD, ApoE, ApoF, ApoH, ApoJ, ApoL types 1 and 2; ApoM, and ApoO, focusing on those with conflicting pathogenicity interpretations (cpi) in dbSNP [14], Ensembl [15], and the relevant PubMed literature. The cpi criterion was used to cluster bio-computational data on this topic. Various computational tools were employed to conduct diverse predictions and calculations, with detailed discussions provided in the subsequent sections. We employ a multifaceted approach that integrates predictions from various computational tools, including SIFT [16], PolyPhen-2 [17], FATHMM [18], SNPs&Go [19], mCSM [20], DynaMut2 [21], MAESTROweb [22], PremPS [23], MutPred2 [24], PhDSNP [25], Camsol [26], Aggrescan3D 2.0 [27] and Frustratometer [28]. The explanation of the scoring system for each tool is provided underneath of each table. Details for each tool are present in Supplementary Material in brief here after.

2.2. Predictions Derived from Sequence Analysis

1. Sorts Tolerant From Intolerant (SIFT): bioinformatics tool that predicts the potential impact of amino acid substitutions on protein function [16]. Advantage of SIFT, it allows the analysis of multiple variants, suitable for high-throughput genetic screening studies. A limitation of SIFT, however, is its dependence on multiple sequence alignments to assess the conservation of amino acid residues. When a protein lacks a sufficient number of homologous sequences, the algorithm may struggle to generate accurate predictions.
2. PolyPhen-2: bioinformatics tool that predicts the possible effects of amino acid substitutions on protein structure and function [17]. It integrates comparative and physical properties of amino acids substitutions disrupting the native conformation of the protein. It gathers relevant information, which improves the robustness of predictions. The limit is the indirect prediction of disease risk.
3. Functional Analysis Through Hidden Markov Models (FATHMM): bioinformatics tool used for functional impact of genetic variants prediction, particularly in coding regions of the genome [18]. The predictive accuracy depends heavily on the quality and size of the training dataset. For variants that are not well-represented in the training data, the predictions may be less reliable.
4. SNPs&GO: web-based tool that utilizes a support vector machine (SVM) to identify deleterious single-nucleotide polymorphisms that lead to amino acid substitutions [19]. The SVM classifier integrates protein sequence data, functional profiles,

and gene ontology (GO) annotations to differentiate between disease-associated and neutral variants. It applies machine learning algorithms trained on various features derived from protein sequence and structure to assess the potential effects of SNPs. A score greater than 0.5 suggests that a given substitution is likely to be pathogenic.

2.3. Predictions Derived from Structure Analysis

1. Mutation Cutoff Scanning Matrix (mCSM): web-based tool used in structural bioinformatics to predict the effects of variants on protein stability [20]. It operates by evaluating single-point amino acid substitutions using a graph-based method. It can predict changes in protein stability. The limit is the indirect prediction of aminoacidic changes might affect the protein's function or interactions, which may be critical in understanding disease mechanisms.
2. Dynamut2: tool developed to predict the effects of amino acid substitutions on protein stability [21]. DynaMut2 employs a combination of Normal Mode Analysis (NMA) and molecular dynamics, providing a more detailed assessment of the effects of mutations on protein stability compared to other tools that rely solely on structural data. While DynaMut2 assesses protein stability, it does not directly predict the functional consequences of mutations, such as changes in enzymatic activity or alterations in protein–protein interactions.
3. MAESTROweb: tool used to discern how variants impact protein stability, specifically focusing on the functional implications of genetic variations concerning human diseases [22]. By combining various computational tools, MAESTROweb provides insights into the consequences of variants on protein stability. While MAESTROweb assesses stability changes, the tool does not directly predict how mutations affect protein function, enzymatic activity, or molecular interactions.
4. PremPS: tool that evaluates the impact of amino acid substitutions in proteins by employing multiple approaches, such as multiple sequence alignment (MSA), protein structure analysis, and a deep learning model [23] providing more accurate predictions compared to methods based solely on sequence or structure. One limitation is that predictions may be less reliable if no experimental protein structure is available, and homologous models are used instead.

2.4. Prediction of Pathogenic Potential

1. MutPred2: web-based tool designed to predict the potential pathogenicity of amino acid substitutions. It does this by assessing the likelihood that a variant impacts protein function and structure [24]. MutPred2 provides insights into possible molecular mechanisms by which mutations can contribute to disease, such as altered protein–protein interactions, post-translational modifications, and structural changes. It evaluates both the structural and functional impact of amino acid substitutions, providing insights into potential disease mechanisms. While it generates hypotheses, experimental validation is necessary to confirm the predicted effects.
2. PhD-SNP: tool designed to predict the impact of single nucleotide polymorphisms (SNPs) on protein function and their association with disease [25]. PhD-SNP is generally accessible through web-based interfaces, facilitating its use by researchers without requiring advanced computational expertise. However, as with many predictive tools, there exists the potential for both false positives and false negatives.

2.5. Profiling of Aggregation Propensity and Residual Frustration Analysis

1. Camsol: tool method begins by using a linear combination of biophysical properties, to determine a solubility score for each individual residue [26]. The protein's overall solubility score, derived from its intrinsic solubility profile, reflects the contributions

of both poorly and highly soluble regions [29]. CamSol applies to diverse proteins, including those in disease and biotechnology, but excludes external factors like protein interactions or post-translational modifications.

2. Aggrescan3D 2.0: web-based tool for predicting aggregation-prone regions, assessing mutation effects, and comparing protein aggregation profiles [27]. It quantifies intrinsic aggregation propensities of amino acids, providing insights into residue-specific contributions [26,27]. While valuable, the model may not fully capture complex interactions or folding dynamics influencing aggregation in biological contexts.
3. Frustratometer: server was used to calculate individual and configurational residual indices for the structure of ApoE. This analysis provides valuable insights into a protein's propensity to aggregate, which is useful for understanding how specific variants within the protein influence its tendency to form aggregates [28]. The frustration index might require careful interpretation; as high frustration does not necessarily correlate with aggregation in all contexts. Experimental validation is needed to confirm predictions.

3. Results

The Apos resulted positive in dbSNPs for pathogenicity were seventeen in total; in particular, ApoB counted twelve results, representing 70% of the total. The ApoC types 2 and 3 gave one result for each, represented in total 12%, and ApoE representing 18% which gave three results (Figure 1). The protein sequences of ApoB, ApoC2, ApoC3, and Pre-ApoE, were extracted via UniProt (Accession ID: P04114, P02655, P02656, P02649, respectively).

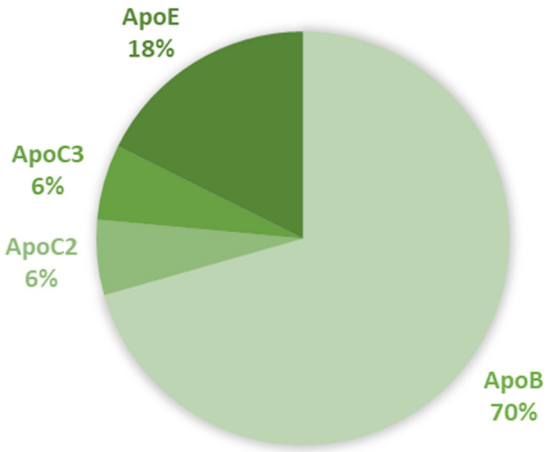

Figure 1. The pie chart displays the distribution of missense SNPs with conflicting interpretations of pathogenicity in dbSNP across apolipoprotein genes. The percentages are color-coded as follows: forest green represents the percentage of SNPs in the Apolipoprotein E (ApoE) gene, emerald green indicates the percentage in the Apolipoprotein C3 (ApoC3) gene, sage green corresponds to the Apolipoprotein C2 (ApoC2) gene, and mint green highlights the Apolipoprotein B (ApoB) gene.

A comprehensive gathering of individual amino acid substitutions within ApoB, ApoC2, ApoC3 and ApoE, was conducted by accessing data from dbSNP and Ensembl. The three-dimensional structure of human ApoB, ApoC2, ApoC3 and pre-ApoE, was retrieved from the Alfaphold. For this latter, the variants reported here are in respect to the mature ApoE, amino acid (aa) replacement leucine (L) to proline (P) in position 46 indicated as L46P, and arginine (R) to cysteine (C) in the 163-position indicated as R163C. List of amino acids replacement resulted with pathogenicity are in the Supplementary Table S1. The functional impact of these mutations was predicted using sequence-based tools, including SIFT, PolyPhen-2, FATHMM, and SNP&GO are presented in Table 1. Among the

mutations analyzed the L46P and R163C in ApoE were predicted to be damaging by three out of four tools. SNPs flagged as deleterious by three out of four sequence-based methods were subjected to further analysis. The prediction score reflects the overall response from the tools, with a value of zero indicative of not significant alterations in sequence, structure, or pathogenicity according to entirely tools employed. Based on these findings, both mutations (L46P and R163C) underwent further bio-computational analysis.

Table 1. Prediction Score using sequence-based tools of mutations, indicated by amino acid (aa) replacement, SNP identity (rs) in the apolipoproteins. In bolt the apolipoproteins with high prediction.

Gene	aa Replacement	SNPs ID (rs)	SIFT Class [1]	PolyPhen-2 [2]	FATHMM [3]	SNPs&GO [4]	Prediction Score
ApoB	T191A	rs1318040190	Tolerated	Benign	Tolerated	Neutral	0/4
ApoB	F299V	rs72653060	Tolerated	Probably Damaging	Tolerated	Neutral	1/4
ApoB	N902K	rs1801700	Deleterious	Possibly Damaging	Tolerated	Neutral	2/4
ApoB	R1128H	rs12713843	Deleterious	Possibly Damaging	Tolerated	Neutral	2/4
ApoB	S1203G	rs78875649	Tolerated	Possibly Damaging	Tolerated	Neutral	1/4
ApoB	G2540V	rs571626569	Tolerated	Possibly Damaging	Tolerated	Neutral	1/4
ApoB	E2566K	rs1801696	Deleterious	Benign	Tolerated	Neutral	1/4
ApoB	N2785H	rs2163204	Deleterious	Benign	Tolerated	Neutral	1/4
ApoB	P2821L	rs72653095	Tolerated	Probably Damaging	Tolerated	Neutral	1/4
ApoB	Y3295H	rs186299244	Tolerated	Probably Damaging	Tolerated	Neutral	1/4
ApoB	R3500W	rs144467873	Tolerated	Probably Damaging	Damaging	Neutral	2/4
ApoB	R4385H	rs533755016	Damaging	Benign	Tolerated	Neutral	1/4
ApoC2	K41T	rs120074114	Tolerated	Benign	Tolerated	Neutral	0/4
ApoC3	A43T	rs147210663	Tolerated	Probably Damaging	Damaging	Neutral	2/4
ApoE	**L46P**	**rs769452**	**Tolerated**	**Probably Damaging**	**Damaging**	**Disease**	**3/4**
ApoE	C130R	rs429358	Tolerated	Benign	Tolerated	Disease	1/4
ApoE	**R163C**	**rs769455**	**Damaging**	**Probably Damaging**	**Tolerated**	**Disease**	**3/4**

[1] SIFT score ranges from 0 to 1. A variant is classified as "damaging" if the score is $\leq$0.05 and "tolerated" if it is >0.05. [2] PolyPhen-2 classifies variants into "benign", "possibly damaging", and "probably damaging." A "possibly damaging" variant may affect protein function, while a "probably damaging" variant has a high likelihood of impacting protein function. [3] FATHMM score ranges from 0 to 1. Variants with a score >0.5 are classified as "damaging", while those with a score $\leq$0.5 are predicted to be "tolerated." [4] SNPs&GO provides a score ranging from 0 to 1. A variant is classified as "disease-related" if the score is $\geq$0.5 and "neutral" if it is <0.5.

The two mutations in ApoE, L46P and R163C were tested using structure-based tools. The structure-based approach involved the followed tools, MAESTROweb, PremPS, mCSM and Dynamut2. They were used in order to evaluate single-point amino acid

substitutions in ApoE, for stability/destability score prediction. The L46P and R163C resulted in destabilizing effects in 3 out of 4 tools and 4 out of 4 tools, respectively (Table 2).

Table 2. Prediction Score using structure-based tools indicated as amino acid (aa) replacement, in ApoE.

aa Replacement	MAESTROweb [1]	PremPS [2]	mCSM [3]	DynaMut2 [4]	Prediction Score
L46P	Destabilizing	Destabilizing	Destabilizing	Stabilizing	3/4
ΔΔG	−0.01 kcal/mol	0.25 kcal/mol	−0.554 kcal/mol	0.33 kcal/mol	
R163C	Destabilizing	Destabilizing	Destabilizing	Destabilizing	4/4
ΔΔG	−0.672 kcal/mol	0.71 kcal/mol	−1.565 kcal/mol	−1.28 kcal/mol	

[1] MAESTROweb score below 0 indicates a predicted destabilization of the protein due to the amino acid substitutions. [2] PremPS positive ΔΔG (kcal/mol) score suggests destabilization, a negative score indicates stabilization. [3] mCSM: ΔΔG (kcal/mol) score below 0 suggests that the mutation is likely to significantly affect protein structure. [4] DynaMut2 positive score indicates that the mutation increases protein stability, while a negative score suggests reduced stability.

Data extracted from DynaMut2 highlighted that in the wild-type structure, L46 forms a hydrogen bond with R43, and several polar interactions with Q42, R43, R50, and G49 (Figure 2, panel A). After substitution, Q42 interacts with P46, and the polar bonds are rearranged among proximal amino acids (Figure 2, panel B). Instead, in the wild type structure R163 is stabilized by polar interactions with K161, L167, and L159, and by hydrophobic interactions with T52 (Figure 2 panel C). While, in the mutated protein, C163 is stabilized by van der Waals forces and polar bonds with L167 and R160 (Figure 2, panel D).

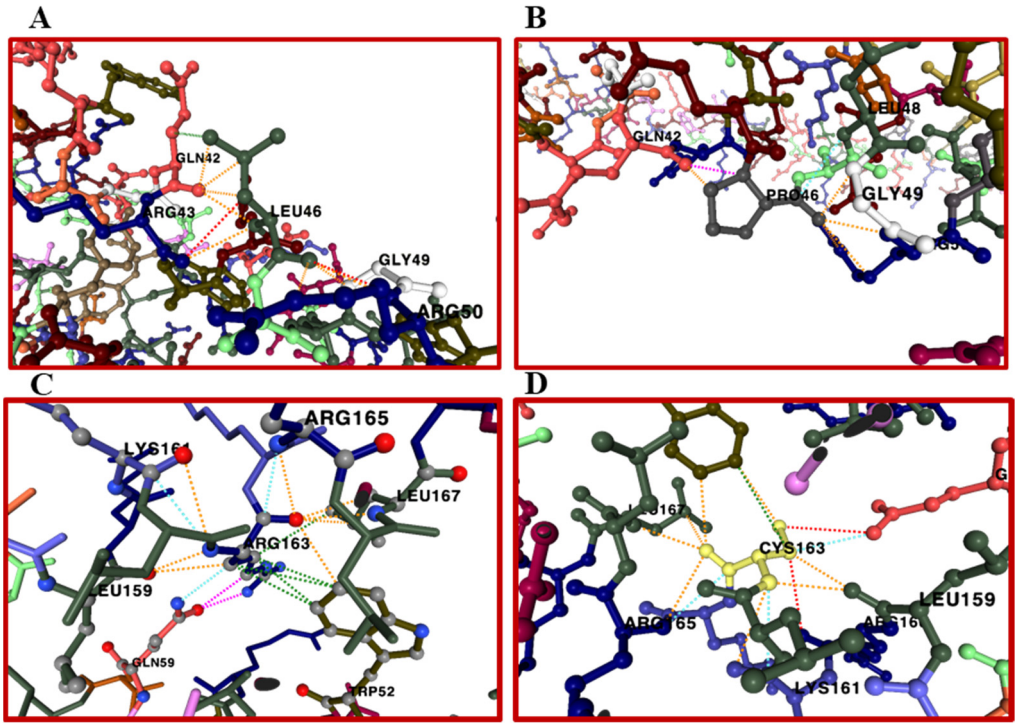

Figure 2. Representation of intramolecular interactions by DynaMut. (**A**) wild type ApoE L46; (**B**) mutated P46; (**C**) wild type ApoE R163; (**D**) mutated C163. Green represents hydrophobic interactions; orange shows polar interactions and pink shows clash. As shown in panel (**B**), the L46P mutation creates a clash between the Pro46 and Gln42 amino acids, resulting in instability. On the other hand, as shown in panel (**A**), in the wild type, Leu46 is stabilized by Gln42 through the formation of a polar bond. As shown in panel (**D**), Cys163 is stabilized by the presence of a greater number of polar bonds compared to the wild type shown in panel (**C**), while hydrophobic interactions and the clash with Gln59 disappear.

Then, using the PhD-SNP and MutPred2 web servers, we were able to classify mutations based on their pathogenicity scores and delineate associated disease phenotypes. The tools reported the mutation L46P, correlated with the onset of several pathologies, with 1 out of 2 tools reported it as pathogenetic, as it is shown in Table 3.

Table 3. Prediction Score of pathogenic potential of mutations in ApoE indicated as amino acid (aa) replacement.

aa Replacement	PhD-SNP [1]	Mutpred [2]	Pathogenicity Score
L46P	**Disease**	Neutral	1/2
R163C	Neutral	Neutral	0/2

[1] PhDSNP classifies mutations into diseases related (desired output set to 0) and neutral polymorphism (desired output set to 1). [2] MutPred2 output includes a general score (g), which represents the probability that the amino acid substitution is pathogenic. When interpreted as a probability, a threshold score of 0.50 indicates potential pathogenicity.

In addition, aggregation propensity of the protein mutants L46P and R163C was assessed in comparison of the ApoE wild type. To analyze the solubility of the protein variants we utilized Camsol [26]. Furthermore, Aggrescan3D 2.0 [27] was used for aggregation propensity. In the Camsol analysis, scores lower than -1 indicate a tendency to promote aggregation, while scores above 1 suggest enhanced solubility. Similarly, Aggrescan3D 2.0 provides a total score, with more negative values corresponding to higher solubility. The results showed that the L46P mutation reduces the aggregation propensity of ApoE, enhancing its solubility. In contrast, the R163C mutation yielded mixed results; one of the tools indicated a decrease in solubility (Table 4).

Table 4. Prediction of solubility and aggregations of mutated ApoE indicated as amino acid (aa) replacement comparing to protein wild type.

aa Replacement	Camsol [1]	Aggrescan3D 2.0 [2]
wild type	1.819611	−401.4035
L46P	1.841586	−408.2955
R163C	1.731944	−405.0661

[1] Camsol scores below −1 indicate a propensity for promoting protein aggregation, whereas scores exceeding +1 are associated with an increased likelihood of aggregation. [2] Aggrescan3D 2.0 generates a total score, where more negative values correlate with higher protein solubility, and vice versa.

Finally, we assess the degree of frustration of the mutated ApoE, L46P and R163C compared to the wild-type. The frustration in ApoE wild type is shown in Figure 3 panel A. with the contact map highlighting substantial changes in frustration patterns for both, L46P (Figure 3 panel B) and R163C (Figure 3 panel C). Specifically, the structure displayed moderate frustration at several sites across the proteins.

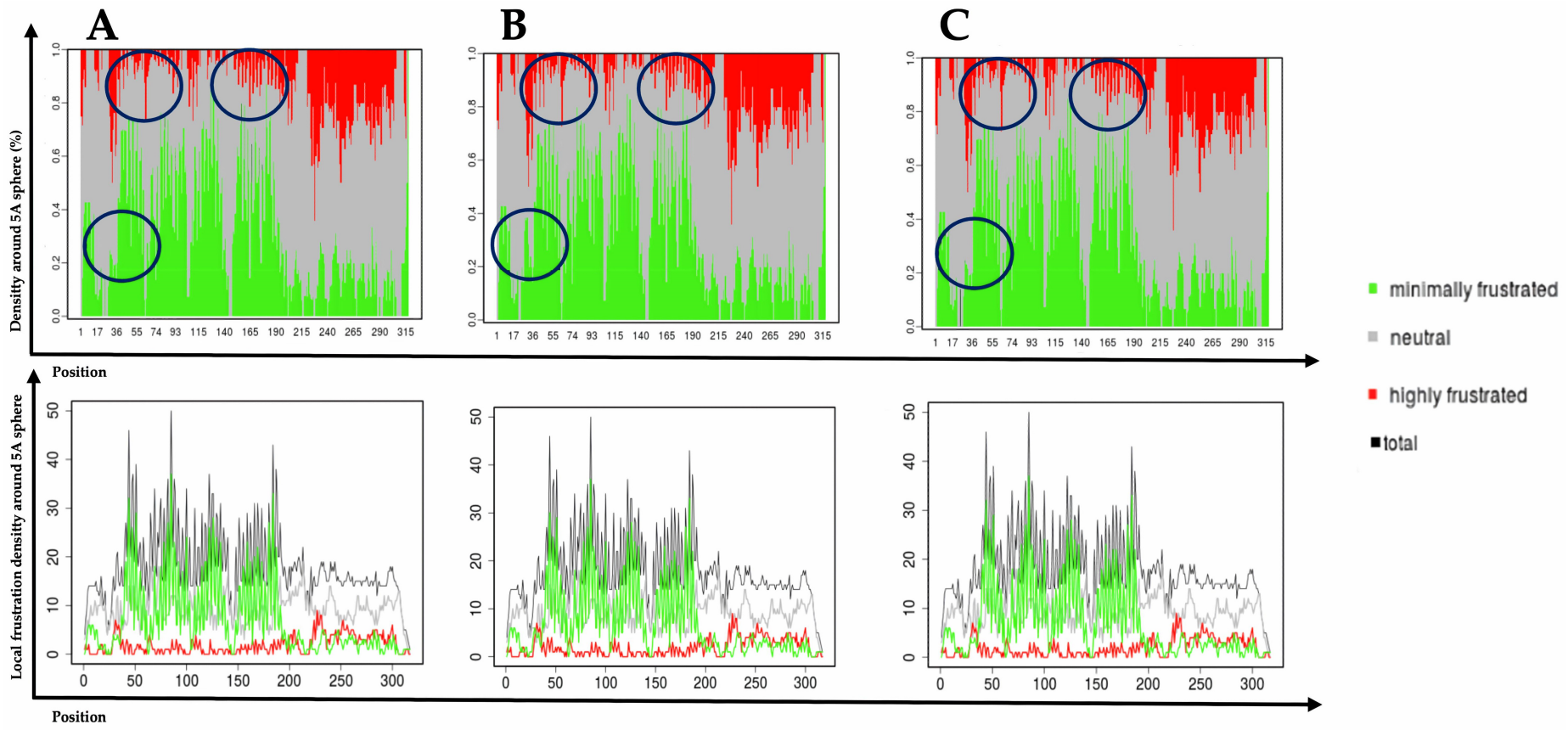

Figure 3. Residual frustration maps for ApoE. (**A**) Wild type; (**B**) L46P; (**C**) R163C. The upper panels shown the density around 5A sphere in percentage. The lower panels shown the local frustration density of 5A sphere as residue-residue contact level. In green is indicated minimally frustrated, in gray neutral, in red highly frustrated and in black (only in the lower panels) the total frustration behavior. The circles highlight the regions of change in terms of molecular frustration as demonstrated by the wild type variant (panel (**A**)), L46P (panel (**B**)), and R163C (panel (**C**)). Changes in areas of frustration may indicate that new energetic configurations could prevail, potentially altering the biological function of the protein.

4. Discussion

APOs are a group of proteins crucial for lipid metabolism, as they are key participants in various transport pathways essential for physiological processes. In humans, several APOs with distinct roles have been identified, including ApoA1 [30], ApoA2 [31], ApoA4 [32], ApoA5 [33], ApoB [34], ApoC1 [35], ApoC2 [36], ApoC3 [37], ApoC4 [38], ApoD [39], ApoE [40], ApoF [41], ApoH [42], ApoJ [43], ApoL1 [44], ApoL2 [45], ApoM [46], ApoO [47]. ApoA and ApoD are integral to the High-Density Lipoprotein (HDL) transport system, with ApoA being the predominant component in plasma. ApoB, on the other hand, is essential for the Low-Density Lipoprotein (LDL) transport system. Similarly, ApoC is primarily associated with Very Low-Density Lipoprotein (VLDL), while ApoE serves as the principal apolipoprotein in chylomicrons. ApoE exists in three common isoforms E2, E3, and E4 exhibiting functional differences and linked to several diseases, including AD [48,49], type III hyperlipoproteinemia [50,51], atherosclerosis [52], cognitive decline [53], telomere shortening [54]. Several molecular mechanisms through which missense mutations contribute to the onset of diseases have been identified [55]. Notably, protein solubility is critical for its functionality, as insoluble regions are susceptible to aggregation [56], which can promote the progression of different diseases [57]. Additionally, protein aggregation is recognized as a risk factor in various disorders [58], as is the destabilization caused by mutations [59]. Dysfunction of APOs has been extensively studied, with specific SNPs being associated with various clinical conditions. In this study, dbSNP was used to identify human SNPs on APOs with conflicting interpretations of pathogenicity. Seventeen SNPs were identified in dbSNPs for ApoB, C2, C3 and E. For this latter, the pre-ApoE sequence was extracted and SNPs in the mature form were shifted of 18-residue, which represent the signal peptide at the beginning of the sequence in the immature form. Here, we referred to the mature ApoE sequence. We identified two mutations with high score confidence SNPs on ApoE, rs769452 (L46P) and rs769455 (R163C). This latter was found to have consistently destabilizing effects on the protein structure, as indicated by all structural prediction tools employed, including mCSM, MAESTROweb, PremPS, and DynaMut2. It was highlighted that in the wild type structure, L46 forms a hydrogen bond with R43 and several polar interactions with Q42, R43, R50, and G49. In the mutant form P46, interacts with Q42, and the polar bonds are rearranged among proximal amino acids to P46. Instead, in the wild-type structure, R163 is stabilized by polar interactions with K161, L167, and L159, and by hydrophobic interactions with T52. In the mutated protein, C163 is stabilized by van der Waals forces and polar bonds with L167 and R160. ApoE, when not associated with lipid, exists as a mixture of monomers, dimers, and tetramers, which is the prevalent species [60] affecting its function in biological micromilieu. Of note, ApoE4 is the prevalent risk factor for LOAD even in the Italian population, as recently analyzed by our group [61]. In a study by Nemergut and co-workers, the authors focused on the peculiar aminoacidic change in ApoE4 in respect to ApoE3, which differs only by the replacement of a cysteine to arginine in 112 position in the pre-ApoE, which correspond to 130 in the mature form. They found that this amino acidic substitution in ApoE4 induces a long-distance (>15 Å) conformational change forming a V-shaped dimeric unit, which, in turn, is more prone to aggregation respect to the ApoE3 structure [62]. To analyze the solubility of the protein variants and their potential aggregation propensity, we utilized Camsol [26] and Aggrescan3D 2.0 [27]. The solubility and aggregation analyses conducted revealed that the L46P mutation results in an increase in solubility, with both tools indicating it as more soluble. In contrast, the R163C mutation is associated with a decrease in solubility according to one of the two tools. The findings emphasize the potential of in silico methodologies to identify high-score confidence deleterious mutations of ApoE, offering a powerful tool for guiding future experimental and clinical research. The computational framework employed in this

study is broadly applicable to other genes and proteins beyond apolipoproteins. It has been successfully applied to proteins involved in autism [63], cancer [64], acute myeloid leukemia [65], Huntington disease [66] demonstrates the potential of this computational platform in identifying mutations with destabilizing effects. Our in silico study highlighted the pathogenicity of both mutations L46P and R163C which, in turn, were found already in a clinical setting, putting in evidence the robustness of integrating multiple bio-computational approaches. As highlighted in the literature, L46P is strongly linked to an increased risk of late-onset Alzheimer's disease (LOAD) [67] and reduced plasma ApoE levels [68], which are further associated with a higher risk of the disease [69,70]. Additionally, the CT genotype of R163C has been associated with the development of obstructive sleep apnea in children [71].

5. Conclusions

The mutations rs769455 (R163C) and rs769452 (L46P) in ApoE studied with the in silico methodologies identified clinically significant genetic variants, highlighting the robustness of the integrated approach. The future direction of the research is to create a multiplex panel with the SNPs identified here in APOE and expand to other proteins to have a panel genetic risk assessment and disease prediction in which ApoE correlates.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biomedinformatics5010016/s1, Table S1: List of amino acid replacements in all APOs, including the mutation name, amino acid substitution notation, wild-type amino acid, mutation position, and the mutated amino acid.

Author Contributions: Conceptualization, E.C. and G.F.S.; methodology, E.C.; software, G.F.S.; validation, E.C. and G.F.S.; formal analysis, G.F.S.; investigation, G.F.S. and E.C.; resources, G.F.S.; data curation, G.F.S.; writing—original draft preparation, G.F.S.; writing—review and editing, G.F.S. and E.C.; visualization, G.F.S. and E.C.; supervision, E.C.; project administration, E.C.; funding acquisition, E.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by PNRR Project: H23C22000370006-T4Y S5G4PP1 to E.C.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Basavaraju, P.; Balasubramani, R.; Kathiresan, D.S.; Devaraj, I.; Babu, K.; Alagarsamy, V.; Puthamohan, V.M. Genetic Regulatory Networks of Apolipoproteins and Associated Medical Risks. *Front. Cardiovasc. Med.* **2022**, *8*, 788852. [CrossRef]
2. Liu, S.; Liu, J.; Weng, R.; Gu, X.; Zhong, Z. Apolipoprotein E gene polymorphism and the risk of cardiovascular disease and type 2 diabetes. *BMC Cardiovasc. Disord.* **2019**, *19*, 213. [CrossRef]
3. Ismail, A.B.; Balcıoğlu, Ö.; Özcem, B.; Ergoren, M.Ç. *APOE* Gene Variation's Impact on Cardiovascular Health: A Case-Control Study. *Biomedicines* **2024**, *12*, 695. [CrossRef] [PubMed]
4. Lozupone, M.; Dibello, V.; Sardone, R.; Castellana, F.; Zupo, R.; Lampignano, L.; Bortone, I.; Daniele, A.; Bellomo, A.; Solfrizzi, V.; et al. The Impact of Apolipoprotein E (*APOE*) Epigenetics on Aging and Sporadic Alzheimer's Disease. *Biology* **2023**, *12*, 1529. [CrossRef]
5. Saraceno, G.F.; Abrego-Guandique, D.M.; Cannataro, R.; Caroleo, M.C.; Cione, E. Machine Learning Approach to Identify Case-Control Studies on *ApoE* Gene Mutations Linked to Alzheimer's Disease in Italy. *BioMedInformatics* **2024**, *4*, 600–622. [CrossRef]

6. Roberts, B.R.; Laffoon, S.B.; Roberts, A.M.; Porter, T.; Fowler, C.; Masters, C.L.; Dratz, E.A.; Laws, S.M. Discovery of a Missense Mutation (Q222K) of the *APOE* Gene from the Australian Imaging, Biomarker and Lifestyle Study. *J. Alzheimer's Dis. Rep.* **2023**, *7*, 165–172. [CrossRef]
7. Perera, S.D.; Hegele, R.A. Genetic variation in apolipoprotein A-V in hypertriglyceridemia. *Curr. Opin. Infect. Dis.* **2024**, *35*, 66–77. [CrossRef]
8. Dedoussis, G.V. Apolipoprotein polymorphisms and familial hypercholesterolemia. *Pharmacogenomics* **2007**, *8*, 1179–1189. [CrossRef]
9. Zhou, F.; Guo, T.; Zhou, L.; Zhou, Y.; Yu, D. Variants in the APOB gene was associated with Ischemic Stroke susceptibility in Chinese Han male population. *Oncotarget* **2017**, *9*, 2249–2254. [CrossRef]
10. Brunzell, J.D.; Davidson, M.; Furberg, C.D.; Goldberg, R.B.; Howard, B.V.; Stein, J.H.; Witztum, J.L. Lipoprotein management in patients with cardiometabolic risk: Consensus statement from the American Diabetes Association and the American College of Cardiology Foundation. *Diabetes Care* **2008**, *31*, 811–822. [CrossRef]
11. Emerging Risk Factors Collaboration; Di Angelantonio, E.; Sarwar, N.; Perry, P.; Kaptoge, S.; Ray, K.K.; Thompson, A.; Wood, A.M.; Lewington, S.; Sattar, N.; et al. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* **2009**, *302*, 1993–2000. [CrossRef]
12. Katsonis, P.; Koire, A.; Wilson, S.J.; Hsu, T.; Lua, R.C.; Wilkins, A.D.; Lichtarge, O. Single nucleotide variations: Biological impact and theoretical interpretation. *Protein Sci.* **2014**, *23*, 1650–1666. [CrossRef] [PubMed]
13. Stenson, P.D.; Mort, M.; Ball, E.V.; Evans, K.; Hayden, M.; Heywood, S.; Hussain, M.; Phillips, A.D.; Cooper, D.N. The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **2017**, *136*, 665–677. [CrossRef]
14. Sherry, S.T.; Ward, M.; Sirotkin, K. dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res.* **1999**, *9*, 677–679. [CrossRef]
15. Harrison, P.W.; Amode, M.R.; Austine-Orimoloye, O.; Azov, A.G.; Barba, M.; Barnes, I.; Becker, A.; Bennett, R.; Berry, A.; Bhai, J.; et al. Ensembl 2024. *Nucleic Acids Res.* **2024**, *52*, D891–D899. [CrossRef]
16. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [CrossRef] [PubMed]
17. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **2013**, *76*, 7.20.1–7.20.41. [CrossRef] [PubMed]
18. Rogers, M.F.; Shihab, H.A.; Mort, M.; Cooper, D.N.; Gaunt, T.R.; Campbell, C. FATHMM-XF: Accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **2018**, *34*, 511–513. [CrossRef]
19. Capriotti, E.; Calabrese, R.; Fariselli, P.; Martelli, P.L.; Altman, R.B.; Casadio, R. WS-SNPs&GO: A web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genom.* **2013**, *14* (Suppl. S3), S6. [CrossRef]
20. Pires, D.E.V.; Ascher, D.B.; Blundell, T.L. mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **2014**, *30*, 335–342. [CrossRef]
21. Rodrigues, C.H.; Pires, D.E.; Ascher, D.B. DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Sci.* **2021**, *30*, 60–69. [CrossRef]
22. Laimer, J.; Hiebl-Flach, J.; Lengauer, D.; Lackner, P. MAESTROweb: A web server for structure-based protein stability prediction. *Bioinformatics* **2016**, *32*, 1414–1416. [CrossRef]
23. Chen, Y.; Lu, H.; Zhang, N.; Zhu, Z.; Wang, S.; Li, M. PremPS: Predicting the impact of missense mutations on protein stability. *PLoS Comput. Biol.* **2020**, *16*, e1008543. [CrossRef]
24. Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K.A.; Lin, G.N.; Nam, H.-J.; Mort, M.; Cooper, D.N.; Sebat, J.; Iakoucheva, L.M.; et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **2020**, *11*, 5918. [CrossRef]
25. Capriotti, E.; Fariselli, P. PhD-SNPg: A webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res.* **2017**, *45*, W247–W252. [CrossRef]
26. Sormanni, P.; Aprile, F.A.; Vendruscolo, M. The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* **2015**, *427*, 478–490. [CrossRef]
27. Zambrano, R.; Jamroz, M.; Szczasiuk, A.; Pujols, J.; Kmiecik, S.; Ventura, S. AGGRESCAN3D (A3D): Server for prediction of aggregation properties of protein structures. *Nucleic Acids Res.* **2015**, *43*, W306–W313. [CrossRef]
28. Parra, R.G.; Schafer, N.P.; Radusky, L.G.; Tsai, M.-Y.; Guzovsky, A.B.; Wolynes, P.G.; Ferreiro, D.U. Protein Frustratometer 2: A tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.* **2016**, *44*, W356–W360. [CrossRef]
29. Sormanni, P.; Amery, L.; Ekizoglou, S.; Vendruscolo, M.; Popovic, B. Rapid and accurate in silico solubility screening of a monoclonal antibody library. *Sci. Rep.* **2017**, *7*, 8200. [CrossRef]
30. Breslow, J.L.; Ross, D.; McPherson, J.; Williams, H.; Kurnit, D.; Nussbaum, A.L.; Karathanasis, S.K.; Zannis, V.I. Isolation and characterization of cDNA clones for human apolipoprotein A-I. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6861–6865. [CrossRef]

31. Brewer, H.B.; Lux, S.E.; Ronan, R.; John, K.M. Amino acid sequence of human apoLp-Gln-II (apoA-II), an apolipoprotein isolated from the high-density lipoprotein complex. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 1304–1308. [CrossRef]
32. Qu, J.; Ko, C.-W.; Tso, P.; Bhargava, A. Apolipoprotein A-IV: A Multifunctional Protein Involved in Protection against Atherosclerosis and Diabetes. *Cells* **2019**, *8*, 319. [CrossRef]
33. Pennacchio, L.A.; Rubin, E.M. *Apolipoprotein A5*, a newly identified gene that affects plasma triglyceride levels in humans and mice. *Arter. Thromb. Vasc. Biol.* **2003**, *23*, 529–534. [CrossRef]
34. Kane, J.P.; Hardman, D.A.; Paulus, H.E. Heterogeneity of apolipoprotein B: Isolation of a new species from human chylomicrons. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 2465–2469. [CrossRef]
35. Fuior, E.V.; Gafencu, A.V. Apolipoprotein C1: Its Pleiotropic Effects in Lipid Metabolism and Beyond. *Int. J. Mol. Sci.* **2019**, *20*, 5939. [CrossRef]
36. Wolska, A.; Dunbar, R.L.; Freeman, L.A.; Ueda, M.; Amar, M.J.; Sviridov, D.O.; Remaley, A.T. Apolipoprotein C-II: New findings related to genetics, biochemistry, and role in triglyceride metabolism. *Atherosclerosis* **2017**, *267*, 49–60. [CrossRef]
37. Vaith, P.; Assmann, G.; Uhlenbruck, G. Characterization of the oligosaccharide side chain of apolipoprotein C-III from human plasma very low density lipoproteins. *Biochim. Biophys. Acta (BBA) Gen. Subj.* **1978**, *541*, 234–240. [CrossRef]
38. Allan, C.M.; Walker, D.; Segrest, J.P.; Taylor, J.M. Identification and characterization of a new human gene (APOC4) in the apolipoprotein E, C-I, and C-II gene locus. *Genomics* **1995**, *28*, 291–300. [CrossRef]
39. McConathy, W.; Alaupovic, P. Isolation and partial characterization of apolipoprotein D: A new protein moiety of the human plasma lipoprotein system. *FEBS Lett.* **1973**, *37*, 178–182. [CrossRef]
40. Utermann, G.; Jaeschke, M.; Menzel, J. Familial hyperlipoproteinemia type III: Deficiency of a specific apolipoprotein (apo E-III) in the very-low-density lipoproteins. *FEBS Lett.* **1975**, *56*, 352–355. [CrossRef]
41. Fidge, N.H.; Nestel, P.J. Identification of apolipoproteins involved in the interaction of human high density lipoprotein3 with receptors on cultured cells. *J. Biol. Chem.* **1985**, *260*, 3570–3575. [CrossRef]
42. Polz, E.; Kostner, G.M. The binding of β_2-glycoprotein-I to human serum lipoproteins: Distribution among density fractions. *FEBS Lett.* **1979**, *102*, 183–186. [CrossRef]
43. Fritz, I.B.; Burdzy, K.; Sétchell, B.; Blaschuk, O. Ram rete testis fluid contains a protein (clusterin) which influences cell-cell interactions in vitro. *Biol. Reprod.* **1983**, *28*, 1173–1188. [CrossRef]
44. Duchateau, P.N.; Pullinger, C.R.; Orellana, R.E.; Kunitake, S.T.; Naya-Vigne, J.; O'Connor, P.M.; Malloy, M.J.; Kane, J.P. Apolipoprotein L, a new human high density lipoprotein apolipoprotein expressed by the pancreas. Identification, cloning, characterization, and plasma distribution of apolipoprotein L. *J. Biol. Chem.* **1997**, *272*, 25576–25582. [CrossRef]
45. Monajemi, H.; Fontijn, R.D.; Pannekoek, H.; Horrevoets, A.J. The apolipoprotein L gene cluster has emerged recently in evolution and is expressed in human vascular tissue. *Genomics* **2002**, *79*, 539–546. [CrossRef]
46. Xu, N.; Dahlbäck, B. A Novel Human Apolipoprotein (apoM). *J. Biol. Chem.* **1999**, *274*, 31286–31290. [CrossRef]
47. Lamant, M.; Smih, F.; Harmancey, R.; Philip-Couderc, P.; Pathak, A.; Roncalli, J.; Galinier, M.; Collet, X.; Massabuau, P.; Senard, J.-M.; et al. ApoO, a novel apolipoprotein, is an original glycoprotein up-regulated by diabetes in human heart. *J. Biol. Chem.* **2006**, *281*, 36289–36302. [CrossRef]
48. Corder, E.H.; Saunders, A.M.; Strittmatter, W.J.; Schmechel, D.E.; Gaskell, P.C.; Small, G.W.; Roses, A.D.; Haines, J.L.; Pericak-Vance, M.A. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **1993**, *261*, 921–923. [CrossRef]
49. Wolk, D.A.; Dickerson, B.C.; Initiative, T.A.D.N.; Weiner, M.; Aiello, M.; Aisen, P.; Albert, M.S.; Alexander, G.; Anderson, H.S.; Anderson, K.; et al. Apolipoprotein E (APOE) genotype has dissociable effects on memory and attentional–executive network function in Alzheimer's disease. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10256–10261. [CrossRef]
50. Mahley, R.W.; Weisgraber, K.H.; Huang, Y. Apolipoprotein E: Structure determines function, from atherosclerosis to Alzheimer's disease to AIDS. *J. Lipid Res.* **2009**, *50*, S183–S188. [CrossRef]
51. Marais, A.D.; Solomon, G.A.E.; Blom, D.J. Dysbetalipoproteinaemia: A mixed hyperlipidaemia of remnant lipoproteins due to mutations in apolipoprotein E. *Crit. Rev. Clin. Lab. Sci.* **2014**, *51*, 46–62. [CrossRef]
52. McNeill, E.; Channon, K.M.; Greaves, D.R. Inflammatory cell recruitment in cardiovascular disease: Murine models and potential clinical applications. *Clin. Sci.* **2010**, *118*, 641–655. [CrossRef]
53. Deary, I.J.; Whiteman, M.C.; Pattie, A.; Starr, J.M.; Hayward, C.; Wright, A.F.; Carothers, A.; Whalley, L.J. Cognitive change and the APOE ε4 allele. *Nature* **2002**, *418*, 932. [CrossRef]
54. Jacobs, E.G.; Kroenke, C.; Lin, J.; Epel, E.S.; Kenna, H.A.; Blackburn, E.H.; Rasgon, N.L. Accelerated cell aging in female APOE-ε4 carriers: Implications for hormone therapy use. *PLoS ONE* **2013**, *8*, e54713. [CrossRef]
55. Stefl, S.; Nishi, H.; Petukh, M.; Panchenko, A.R.; Alexov, E. Molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol.* **2013**, *425*, 3919–3936. [CrossRef]
56. Bhardwaj, T.; Giri, R. Potential of ADAM 17 Signal Peptide to Form Amyloid Aggregates in Vitro. *ACS Chem. Neurosci.* **2023**, *14*, 3818–3825. [CrossRef]

57. Invernizzi, G.; Papaleo, E.; Sabate, R.; Ventura, S. Protein aggregation: Mechanisms and functional consequences. *Int. J. Biochem. Cell Biol.* **2012**, *44*, 1541–1554. [CrossRef]
58. Pan, Q.; Parra, G.B.; Myung, Y.; Portelli, S.; Nguyen, T.B.; Ascher, D.B. AlzDiscovery: A computational tool to identify Alzheimer's disease-causing missense mutations using protein structure information. *Protein Sci.* **2024**, *33*, e5147. [CrossRef]
59. Redler, R.L.; Das, J.; Diaz, J.R.; Dokholyan, N.V. Protein Destabilization as a Common Factor in Diverse Inherited Disorders. *J. Mol. Evol.* **2016**, *82*, 11–16. [CrossRef]
60. Garai, K.; Frieden, C. The association−dissociation behavior of the ApoE proteins: Kinetic and equilibrium studies. *Biochemistry* **2010**, *49*, 9533–9541. [CrossRef]
61. Abrego-Guandique, D.M.; Saraceno, G.F.; Cannataro, R.; de Burnside, M.M.; Caroleo, M.C.; Cione, E. Apolipoprotein E and Alzheimer's Disease in Italian Population: Systematic Review and Meta-Analysis. *Brain Sci.* **2024**, *14*, 908. [CrossRef]
62. Nemergut, M.; Marques, S.M.; Uhrik, L.; Vanova, T.; Nezvedova, M.; Gadara, D.C.; Jha, D.; Tulis, J.; Novakova, V.; Planas-Iglesias, J.; et al. Domino-like effect of C112R mutation on ApoE4 aggregation and its reduction by Alzheimer's Disease drug candidate. *Mol. Neurodegener.* **2023**, *18*, 38. [CrossRef]
63. Owji, H.; Eslami, M.; Nezafat, N.; Ghasemi, Y. In Silico Elucidation of Deleterious Non-synonymous SNPs in SHANK3, the Autism Spectrum Disorder Gene. *J. Mol. Neurosci.* **2020**, *70*, 1649–1667. [CrossRef]
64. Petrosino, M.; Novak, L.; Pasquo, A.; Chiaraluce, R.; Turina, P.; Capriotti, E.; Consalvi, V. Analysis and Interpretation of the Impact of Missense Variants in Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 5416. [CrossRef]
65. Mustafa, M.I.; Mohammed, Z.O.; Murshed, N.S.; Elfadol, N.M.; Abdelmoneim, A.H.; Hassan, M.A. In Silico Genetics Revealing 5 Mutations in *CEBPA* Gene Associated with Acute Myeloid Leukemia. *Cancer Inform.* **2019**, *18*, 1176935119870817. [CrossRef]
66. Shafie, A.; Ashour, A.A.; Anjum, F.; Shamsi, A.; Hassan, I. Elucidating the Impact of Deleterious Mutations on IGHG1 and Their Association with Huntington's Disease. *J. Pers. Med.* **2024**, *14*, 380. [CrossRef]
67. Medway, C.W.; Abdul-Hay, S.; Mims, T.; Ma, L.; Bisceglio, G.; Zou, F.; Pankratz, S.; Sando, S.B.; Aasly, J.O.; Barcikowska, M.; et al. ApoE variant p.V236E is associated with markedly reduced risk of Alzheimer's disease. *Mol. Neurodegener.* **2014**, *9*, 11. [CrossRef]
68. Rasmussen, K.L.; Tybjærg-Hansen, A.; Nordestgaard, B.G.; Frikke-Schmidt, R. *APOE* and dementia—resequencing and genotyping in 105,597 individuals. *Alzheimer's Dement.* **2020**, *16*, 1624–1637. [CrossRef]
69. Rasmussen, K.L.; Tybjærg-Hansen, A.; Nordestgaard, B.G.; Frikke-Schmidt, R. Plasma levels of apolipoprotein E and risk of dementia in the general population. *Ann. Neurol.* **2015**, *77*, 301–311. [CrossRef]
70. Rasmussen, K.L.; Tybjærg-Hansen, A.; Nordestgaard, B.G.; Frikke-Schmidt, R. Plasma apolipoprotein E levels and risk of dementia: A Mendelian randomization study of 106,562 individuals. *Alzheimer's Dement.* **2018**, *14*, 71–80. [CrossRef]
71. Kalra, M.; Pal, P.; Kaushal, R.; Amin, R.S.; Dolan, L.M.; Fitz, K.; Kumar, S.; Sheng, X.; Guha, S.; Mallik, J.; et al. Association of ApoE genetic variants with obstructive sleep apnea in children. *Sleep Med.* **2008**, *9*, 260–265. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.