



## Article

# Exome-Wide Association Study Identified Clusters of Pleiotropic Genetic Associations with Alzheimer's Disease and Thirteen Cardiovascular Traits

Yury Loika <sup>\*</sup> , Elena Loiko, Irina Culminskaya and Alexander M. Kulminski 

Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC 27708, USA; elena.loiko@duke.edu (E.L.); irina.culminskaya@duke.edu (I.C.); alexander.kulminski@duke.edu (A.M.K.)

<sup>\*</sup> Correspondence: yury.loika@duke.edu

**Abstract:** Alzheimer's disease (AD) and cardiovascular traits might share underlying causes. We sought to identify clusters of cardiovascular traits that share genetic factors with AD. We conducted a univariate exome-wide association study and pair-wise pleiotropic analysis focused on AD and 16 cardiovascular traits—6 diseases and 10 cardio-metabolic risk factors—for 188,260 UK biobank participants. Our analysis pinpointed nine genetic markers in the *APOE* gene region and four loci mapped to the *CDK11*, *OBP2B*, *TPM1*, and *SMARCA4* genes, which demonstrated associations with AD at  $p \leq 5 \times 10^{-4}$  and pleiotropic associations at  $p \leq 5 \times 10^{-8}$ . Using hierarchical cluster analysis, we grouped the phenotypes from these pleiotropic associations into seven clusters. Lipids were divided into three clusters: low-density lipoprotein and total cholesterol, high-density lipoprotein cholesterol, and triglycerides. This split might differentiate the lipid-related mechanisms of AD. The clustering of body mass index (BMI) with weight but not height indicates that weight defines BMI-AD pleiotropy. The remaining two clusters included (i) coronary heart disease and myocardial infarction; and (ii) hypertension, diabetes mellitus (DM), systolic and diastolic blood pressure. We found that all AD protective alleles were associated with larger weight and higher DM risk. Three of the four (75%) clusters of traits, which were significantly correlated with AD, demonstrated antagonistic genetic heterogeneity, characterized by different directions of the genetic associations and trait correlations. Our findings suggest that shared genetic factors between AD and cardiovascular traits mostly affect them in an antagonistic manner.

**Keywords:** exome-wide association study; Alzheimer's disease; pleiotropy; hierarchical cluster analysis; aging; cardio-metabolic traits; coronary heart disease; myocardial infarction; stroke; heart failure



**Citation:** Loika, Y.; Loiko, E.; Culminskaya, I.; Kulminski, A.M. Exome-Wide Association Study Identified Clusters of Pleiotropic Genetic Associations with Alzheimer's Disease and Thirteen Cardiovascular Traits. *Genes* **2023**, *14*, 1834. <https://doi.org/10.3390/genes14101834>

Academic Editor: Claudia Ricci

Received: 29 August 2023

Revised: 18 September 2023

Accepted: 20 September 2023

Published: 22 September 2023



**Copyright:** © 2023 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

Alzheimer's disease (AD) is an aging-related neurodegenerative disorder, with few interventions available to ameliorate its symptoms. As the number of AD-affected Americans is expected to increase rapidly, reaching about 12.7 million cases by 2050 [1], identifying the causes of Alzheimer's disease would be beneficial for its prevention and treatment. Several hypotheses have been proposed to explain AD pathology, such as amyloid beta ( $A\beta$ ) cascade, tau pathology, calcium, inflammatory, cholinergic, and oxidative stress hypotheses [2]. It is likely that many biological processes and pathways related to these hypotheses contribute to the initiation and progression of AD. In addition to the buildup of  $A\beta$  plaques and neurofibrillary tangles, the loss of neuronal synapse junctions, neuronal cell death, and brain atrophy can contribute to cognitive impairment underlying AD.

At this time, none of the aforementioned hypotheses alone have succeeded in explaining AD pathology. The amyloid beta hypothesis has been thoroughly investigated in the clinical treatment of AD, with a focus on targeting  $A\beta$  plaques. Notably, two drugs, lecanemab and aducanumab, have received approval from the FDA through the accelerated approval mechanism. Additionally, a third drug, donanemab, has shown promising results

in clinical trials [3], but there is still no cure for AD at this time. This means that there are other processes accompanying the disease and contributing to its progression and severity. This includes the breakdown of the blood–brain barrier (BBB), the immune system response, reduced cerebral blood flow, altered adult neurogenesis and energy metabolism, cell membranes and the cell cycle, the cytoskeleton, and lipid and protein metabolism [4,5]. These processes can develop independently and/or be initiated by other age-related traits or events. The correlation of AD with these other traits may indicate common biological processes underlying them. Therefore, studying the patterns in the pleiotropic associations of AD and the other traits can facilitate identifying and understanding the biological mechanisms underlying AD pathogenesis.

Genome- and exome-wide association studies (GWAS and EWAS, respectively) are comprehensive tools designed to identify genes and related biological mechanisms associated with different traits. While conventional GWAS and EWAS focus on identifying genetic associations with single traits, pleiotropic GWAS and EWAS take a broader approach. They aim to simultaneously assess genetic associations with multiple traits and identify shared genetic components that contribute to the pleiotropy observed in epidemiological studies. Previous epidemiological studies that considered AD and other related traits helped identify several cardiovascular factors conferring AD risk. For instance, increased levels of LDL-C and TC are considered risk factors for AD [6,7], while increased levels of HDL-C were highlighted as protective [8]. Pleiotropic GWAS/EWAS provided an additional opportunity to identify several genetic components and respective biological mechanisms shared by those multiple traits [9,10].

Certain risk factors can form clusters, i.e., they increase or decrease AD risks simultaneously by acting in additive or synergistic manners. Several studies have already demonstrated the effects of such clusters [11,12]. For example, the clustering of vascular risk factors, such as midlife obesity, high SBP, and high total cholesterol levels, increased the risk of dementia and AD in an additive manner [12]. Clusters of hypertension and heart disease, diabetes, and current smoking also increased AD risk [11].

Many of the previous analyses of clustering of neurodegenerative conditions, including AD, used individual-level data for phenotypes as a distance measure [13–17]. In such types of cluster analyses, strongly correlated phenotypes usually cluster together. This approach is not quite efficient, however, in cases when the underlying biological mechanisms and related genetic components are not major contributors to the phenotypes' correlations. In this study, we applied an approach in which the statistics of univariate and pleiotropic genetic associations were used for identifying clusters of AD and its risk factors. We believe that this approach can more accurately characterize complex roles of genetic and non-genetic factors in AD pathogenesis [18–21], including complex interactions, which can be attributed to different biological mechanisms driving antagonistic genetic heterogeneity [22,23]. Such heterogeneity was shown to be widespread in the genetics of lipid traits [24]. This phenomenon was also observed in pleiotropic associations with AD and some traits related to educational attainment and cardiovascular risk factors [25–27].

The goal of this study was to gain insights into pleiotropic predisposition to cardiovascular and AD risk factors and demonstrate complex clustering patterns of AD with 16 risk factors. We hypothesized that using summary statistics of pleiotropic associations among AD and related phenotypes (risk factors) could capture clusters of phenotypes based on the similarities among the genetic effects, rather than solely on the direct correlations among these phenotypes. This approach may facilitate the discovery of shared biological pathways. An additional goal was to investigate how antagonistic genetic heterogeneity shapes the clustering of AD-related phenotypes. For this purpose, we carried out a pair-wise pleiotropic exome-wide association study (~260 K common genetic variants) on predisposition to AD and on each of the 16 selected traits in a sample of ~190 K individuals from the UK Biobank. The analysis included six qualitative traits (coronary heart disease (CHD), diabetes mellitus (DM), stroke, myocardial infarction (MI), heart failure (HF), and hypertension (HT)), and ten other cardio-metabolic risk factors (blood glucose (BG), BMI, height, weight, four lipid traits

(LDL-C, HDL-C, TG, and TC), and systolic (SBP) and diastolic (DBP) blood pressure). Fisher's method was used to test the pleiotropic associations, and hierarchical cluster analysis was used to identify clusters of AD risk factors with similar genetic components.

## 2. Materials and Methods

### 2.1. Accession Numbers

This research has been conducted using data from the UK Biobank, a major biomedical database (<http://www.ukbiobank.ac.uk/>, accessed on 15 March 2021).

### 2.2. Study Cohorts

Data from the UK Biobank (UKB) [28,29] on individuals of Caucasian ancestry, men and women combined, were considered in the analyses. Table 1 provides the basic demographic information of the genotyped participants. After quality control of genetic and phenotypic information, 188,260 individuals remained in the subsequent analyses.

**Table 1.** Basic demographic information of the UK biobank participants with whole exome genotyping information.

AD Status	PAR	N	Women	BC Range	LS	Age	Mortality			
	VALUE (SD)	188,260	103,736 (55.10)	1936–1970	65.21 (8.0)	56.70 (8.03)	899 (0.48)			
<b>Quantitative traits</b>										
	AD	DM	HT	CHD	MI	STROKE	HF			
Cases	361	53 (15%)	189 (52%)	89 (25%)	37 (10%)	30 (8%)	20 (6%)			
Controls	187,899	11,015 (6%)	45,349 (24%)	16,327 (9%)	5834 (3%)	4589 (2%)	3236 (2%)			
<i>p</i> -value		$1.33 \times 10^{-9}$	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$	$3.90 \times 10^{-10}$	$9.97 \times 10^{-9}$	$6.867 \times 10^{-6}$			
<b>Quantitative traits</b>										
	BG (mg/dL)	BMI (kg/m <sup>2</sup> )	Height (cm)	Weight (kg)	SBP (mmHg)	DBP (mmHg)	HDL-C (mg/dL)	LDL-C (mg/dL)	TC (mg/dL)	TG (mg/dL)
Cases	93.36 (19.93)	27.01 (4.90)	167.18 (9.17)	75.62 (15.40)	146.18 (19.58)	81.79 (10.96)	57.77 (16.30)	133.77 (35.72)	216.97 (47.43)	149.26 (78.77)
Controls	92.13 (21.19)	27.34 (4.73)	168.67 (9.24)	78.03 (15.83)	139.78 (19.58)	82.13 (10.65)	56.49 (14.82)	138.07 (33.44)	221.19 (43.99)	154.26 (89.82)
<i>p</i> -value	$5.89 \times 10^{-1}$	$5.05 \times 10^{-1}$	$1.61 \times 10^{-1}$	$1.72 \times 10^{-1}$	$3.46 \times 10^{-2}$	$9.04 \times 10^{-1}$	$4.77 \times 10^{-1}$	$2.68 \times 10^{-1}$	$3.96 \times 10^{-1}$	$5.54 \times 10^{-1}$

PAR = the name of characteristics/parameter; SD = standard deviation. N = sample size; Women = number of women (%); BC Range = range of birth years; LS = life span in years; age = mean age measured in years at the selected examination of quantitative traits. *p*-values for qualitative and quantitative phenotypes were assessed by using the exact Fisher and Wald test, respectively. AD = Alzheimer's disease; DM = diabetes mellitus; HT = hypertension; CHD = coronary heart disease; MI = myocardial infarction; STROKE = stroke; HF = heart failure. BG = blood glucose; BMI = body mass index; SBP = systolic blood pressure; DBP = diastolic blood pressure; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; TC = total cholesterol; TG = triglycerides.

### 2.3. Genotypes

Exome sequencing data for ~200 K individuals were available from the UK Biobank Exome Sequencing Consortium [30]. In the analyses, we included directly genotyped markers (no imputation was performed) with minor allele frequency (MAF) greater than 0.5%, which resulted in 258,684 (~250 K) genetic variants. A total of 213,935 SNPs had a missing call rate better than 5%. For these SNPs, all individuals had a missing call rate better than 5%. Only these SNPs were considered for reporting the results of this study. We did not apply the Hardy–Weinberg equilibrium test at this stage because negligible deviation from Hardy–Weinberg equilibrium was still highly significant in the large UKB sample.

### 2.4. Phenotypes

We considered incident cases of AD and six other diseases, including DM, HT, CHD, MI, stroke, and HF (Table 1). All cases were defined by the UK Biobank based on the ICD-9

and ICD-10 codes. AD status was defined by using codes F00 and G30 according to the ICD-10 classification.

Ten quantitative phenotypes, BG (mg/dL), BMI (kg/m<sup>2</sup>), height (m), weight (kg), SBP (mmHg), DBP (mmHg), HDL-C (mg/dL), TG (mg/dL), LDL-C (mg/dL), and TC (mg/dL) were considered. Because longitudinal information (multiple measurements) was available for a small number of participants, we selected measurements based on the first examination for which information was available (Table 1) (see also Section 2.7). We found significant differences between the AD cases and controls in the qualitative phenotypes (assessed using the exact Fisher test) and the mean values for SBP (evaluated using the Wald test).

### 2.5. Correlations among Phenotypes

Information on the Pearson correlation coefficients between AD and each of the considered quantitative and qualitative phenotypes and their significance is given in Table 2. We observed a small positive significant correlation between AD and each of the six diseases. The largest and most significant correlation was between AD and HT ( $r = 0.029$ ,  $p = 2.41 \times 10^{-23}$ ). Among the quantitative traits, only weight, height, and SBP demonstrated small significant correlations with AD (Table 2).

**Table 2.** Pearson correlation coefficient between Alzheimer’s disease and each of the considered phenotypes in the UK Biobank based on individual-level data and summary statistics of genetic associations.

Phenotype	Individual-Level Data		Summary Statistics	
	r	P	r	P
HT *	0.0290	$2.41 \times 10^{-23}$	0.0352	$<2.2 \times 10^{-16}$
CHD *	0.0248	$1.82 \times 10^{-17}$	0.0145	$2.15 \times 10^{-11}$
MI *	0.0180	$6.52 \times 10^{-10}$	−0.0169	$4.67 \times 10^{-15}$
STROKE *	0.0166	$1.19 \times 10^{-8}$	−0.0174	$8.10 \times 10^{-16}$
DM *	0.0164	$1.76 \times 10^{-8}$	0.0365	$<2.2 \times 10^{-16}$
HF *	0.0128	$1.08 \times 10^{-5}$	−0.0358	$<2.2 \times 10^{-16}$
BG	0.0025	$3.82 \times 10^{-1}$	−0.0069	$1.55 \times 10^{-3}$
BMI	−0.0031	$2.82 \times 10^{-1}$	0.0514	$<2.2 \times 10^{-16}$
Height *	−0.0070	$1.61 \times 10^{-2}$	−0.0700	$<2.2 \times 10^{-16}$
Weight *	−0.0066	$2.30 \times 10^{-2}$	0.0092	$2.03 \times 10^{-5}$
SBP *	0.0143	$9.34 \times 10^{-7}$	−0.0082	$1.41 \times 10^{-4}$
DBP	−0.0014	$6.35 \times 10^{-1}$	0.0164	$2.87 \times 10^{-14}$
HDL-C	0.0038	$1.96 \times 10^{-1}$	−0.0074	$5.84 \times 10^{-4}$
LDL-C	−0.0057	$5.17 \times 10^{-2}$	−0.0029	$1.87 \times 10^{-1}$
TC	−0.0042	$1.48 \times 10^{-1}$	−0.0086	$7.62 \times 10^{-5}$
TG	−0.0024	$4.00 \times 10^{-1}$	−0.0078	$3.41 \times 10^{-4}$

r = Pearson correlation coefficient; P = *p*-values; \* = significant correlation based on individual-level data. DM = diabetes mellitus; HT = hypertension; CHD = coronary heart disease; MI = myocardial infarction; STROKE = stroke; HF = heart failure. BG = blood glucose; BMI = body mass index; SBP = systolic blood pressure; DBP = diastolic blood pressure; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; TC = total cholesterol; TG = triglycerides.

Figure S1 and Table S1 display the Pearson correlation coefficients among the various traits under consideration. Most correlations were statistically significant, with the exception of those involving stroke–height and the ones noted above for AD. Among the observed correlations, the highest were between LDL-C and TC ( $r = 0.95$ ), CHD and MI ( $r = 0.58$ ), SBP and DBP ( $r = 0.67$ ), weight and BMI ( $r = 0.83$ ), and weight and height ( $r = 0.54$ ). Conversely, the most pronounced negative correlation was between HDL-C and TG ( $r = -0.44$ ), as well as HDL-C and weight ( $r = -0.46$ ).

## 2.6. Correlations among Summary Statistics

Table 2 also includes information on the Pearson correlation coefficients that were obtained by using summary statistics of the genetic associations with AD and each of the considered quantitative and qualitative phenotypes. Only the correlation coefficient between the summary statistics for AD and LDL-C was statistically non-significant. The correlation coefficients of the summary statistics for the other traits with AD were statistically significant. Moreover, for five traits (MI, stroke, HF, weight, and SBP), the correlation based on individual-level data and on summary statistics demonstrated opposite directions and were statistically significant. This demonstrates that correlations between the phenotypes at an individual level can be shaped by other factors in addition to genetic component(s).

## 2.7. Statistical Analyses

**The univariate unconditional analysis**—i.e., the univariate exome-wide association study (EWAS)—of the associations of each SNP with each quantitative phenotype or disease was performed using a linear or logistic regression model, as implemented in *plink 2.0* software. An additive genetic model was considered with the minor allele as an effect allele. All models in all analyses were adjusted for sex and age at the selected examination. Therefore, our EWAS provided 17 summary statistics (effect size, standard error, and *p*-value) for the estimates of the associations with seven diseases and ten continuous traits for each SNP.

**Pleiotropic meta-analysis** was performed for the pairs of traits. We used an AD-centric approach when each pair included AD and one of the other traits. Given the small correlations among the traits (Table 2), Fisher's method [31] was used for pleiotropic analysis. It combines *p*-values across phenotypes, disregarding the effect directions and correlations among them, and addresses the issue of multiple testing by increasing the number of degrees of freedom. Therefore, 16 pleiotropic meta-analyses were conducted.

## 2.8. Pleiotropic Associations

As Fisher's method corrects pleiotropic *p*-values for testing multiple phenotypes, the traditional genome-wide (GW) level of significance,  $p_{GW} = 5 \times 10^{-8}$ , was used to obtain conclusions about the significance of the pleiotropic effects. Here, we report GW significant pleiotropic SNPs when at least one of 16 pleiotropic tests attained GW significance,  $p < p_{GW}$ , and the EWAS univariate associations with these traits demonstrated significant association at the levels of  $p < 5 \times 10^{-4}$  for AD and  $p < 5 \times 10^{-2}$  for the other traits. As this analysis was AD-centric, we applied a more stringent threshold for AD associations and a less stringent threshold for each of the other phenotypes.

## 2.9. Index SNPs and Gene Mapping

The NCBI dbSNP database and variant effect predictor from Ensembl (assembly GRCh38.p13) were used for mapping SNPs to genes. Outside of the *APOE* gene region, one index SNP was selected per locus. Multiple genes were reported when several genes overlapped. The index SNPs were selected based on the most significant association (smallest *p*-values) within each genetic locus. Furthermore, the selection of index SNPs took into consideration the proximity of the MAF from the UKB exome chip to the MAF found in the UK10 K TWINS and/or 1000 G reference datasets, as the MAF for some SNPs substantially differed. SNPs that met the Hardy–Weinberg threshold  $p\text{-value} = 1 \times 10^{-40}$ , were retained. We used a small *p*-value for the Hardy–Weinberg cut-off because even very small deviation from Hardy–Weinberg equilibrium was highly significant in the large UKB sample.

## 2.10. Cluster Analysis

Hierarchical cluster analysis can be applied if a distance measure is defined. Previous studies [11–17] used individual-level data of phenotype measurements as a distance measure, which assesses the similarity of phenotypes, including their genetic effects, effects

of exogenous exposures, and their interaction. In this study, we proposed using genetic associations with phenotypes as a distance measure. This approach is more appropriate for gaining insight into the similarities related to genetic components rather than exogenous exposures, considering that the contributions of genetic effects and exogenous exposures on the correlations among the phenotypes can be substantially different (see Table 2).

In this study, SNP–phenotype clusters were identified, leveraging the results of our univariate and AD-centric pair-wise pleiotropic analyses. First, we identified patterns characterized by the relative directions of the associations of the same alleles with AD and each of the other phenotypes and the significance of the pleiotropic associations. The relative directions were defined by the sign of a product of the effect sizes of the associations of SNPs with AD and the other phenotype in a pair. For instance, if the associations of SNPs with these phenotypes were of the same (opposite) directions, such SNPs had the same (different) pattern.

These patterns can be subdivided further into clusters based on the  $p$ -values of the associations of SNPs with phenotypes. To improve the resolution of the cluster analysis, we used  $p$ -values from the univariate analyses of non-AD phenotypes and selected pleiotropic SNPs for which the significance of their associations with AD was  $p < 5 \times 10^{-4}$ . We further categorized minus-log-transformed  $p$ -values from the univariate associations of non-AD phenotypes,  $-\log(p_{uni})$ , into five categories, denoted as 0 for  $p > 0.05$ , 1 for  $5 \times 10^{-4} \leq p < 5 \times 10^{-2}$ , 2 for  $5 \times 10^{-6} \leq p < 5 \times 10^{-4}$ , 3 for  $5 \times 10^{-8} \leq p < 5 \times 10^{-6}$ , and 4 for  $p < 5 \times 10^{-8}$ . This procedure created an  $A_{ij}$  matrix for the  $i$ th SNP and  $j$ th phenotype, in which the sign of each element was determined by the relative directions of the SNP associations with AD and a respective trait, and its magnitude was defined by the aforementioned categories.

Two approaches were used for cluster analysis. Both approaches utilized hierarchical cluster analysis with Euclidian measure and Ward's method. Under this method, associations are clustered by minimizing the distance measure inside the clusters and maximizing this measure between clusters. Our main approach was applied to matrix  $A_{ij}$  as implemented by the *hclust* function in R. This approach estimates clusters based on the similarity of significances and sign-directions of the SNP–phenotype associations. The ad-hoc height cut-off threshold for selecting the clusters was equal to five. This analysis was validated using an approach based on Pearson's correlation as a distance measure. It was implemented using the *heatmap* function in R package *ecodist*. This approach estimates clusters based on the collinearity of phenotypes as vectors of the associations with selected SNPs (collinearity of columns of matrix  $A_{ij}$ ) and on the collinearity of the associations of selected SNPs with the considered phenotypes (collinearity of rows of matrix  $A_{ij}$ ).

### 3. Results

Figure 1 represents a flowchart of the analyses performed in this study.

#### 3.1. Univariate Associations from EWAS

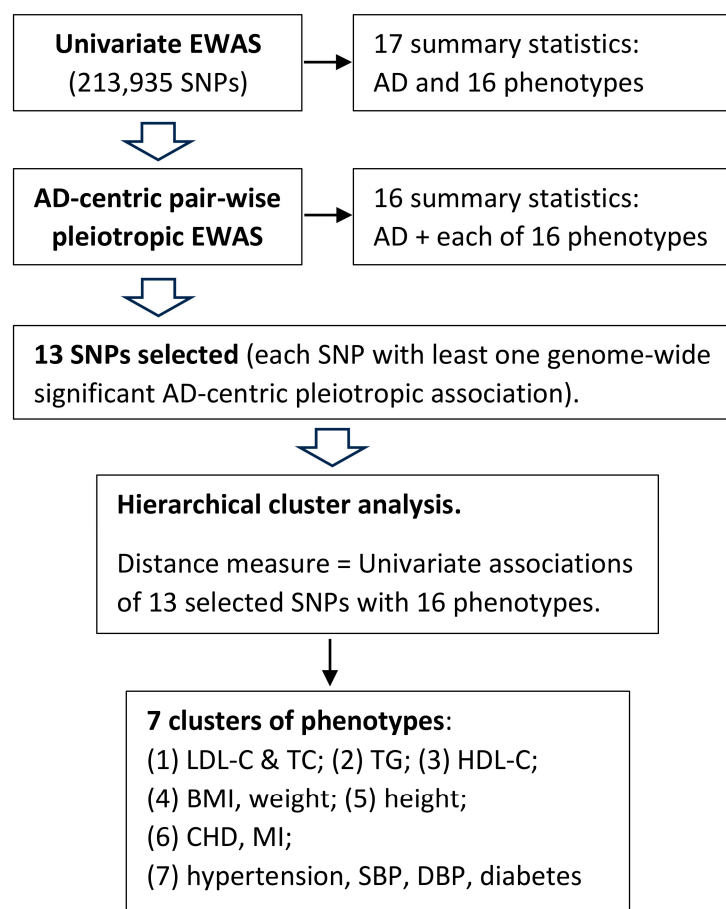
A univariate EWAS was performed for seven diseases (DM, HT, CHD, MI, HF, stroke, and AD) and ten quantitative phenotypes (BG, BMI, weight, height, SBP, DBP, HDL-C, TG, LDL-C, and TC) in a sample of 188,260 UK biobank participants of European ancestry, for men and women combined (see Section 2).

For the diseases, the only GW significant associations were observed with AD in the *APOE* gene cluster on chromosome 19 for rs112849259 and rs741780 *TOMM40* SNPs and rs440446 and rs429358 *APOE* SNPs (Tables 3 and S2, Supplementary Materials). Also, for SNPs in the *APOE* gene cluster, we identified GW significant associations with at least one of the lipid traits: HDL-C (three SNPs), LDL-C (six SNPs), TG (five SNPs), and TC (six SNPs). In total, this cluster of genes harbored 24 GW significant univariate SNP–trait associations (Table S2).

**Table 3.** Associations with risk of Alzheimer’s disease: SNPs from 4 genetic loci and *APOE* gene region.

N	Gene(s) <sup>1</sup>	SNP <sup>2</sup>	Chr	Location, Base Pairs GRCh38	Ref /Alt	MAF obs (%)	P <sub>HWE</sub>	Function	Beta	SE	P	Gene GRASP	SNP GRASP	r <sup>2</sup>	P <sub>GRASP</sub>	PMID
1	CDK11B (MMP23B)	rs28688376	1	1,637,577	T/c	26.1	$6 \times 10^{-36}$	intron	0.298	0.079	$1.56 \times 10^{-4}$					
2	OBP2B	rs11244035	9	133,205,932	C/t	10.2	$2 \times 10^{-30}$	missense	0.393	0.109	$3.23 \times 10^{-4}$	ABO	rs8176694	0.02	$1.3 \times 10^{-2}$	20061627
3	TPM1	rs4775613	15	63,056,897	A/g	42.8	$3 \times 10^{-2}$	5'UTR	−0.334	0.078	$1.99 \times 10^{-5}$	RAB8B	rs10519190	0.00	$2.2 \times 10^{-4}$	17998437
4	SMARCA4 APOE	rs28997580	19 19	11,013,062	C/t	0.9	$3 \times 10^{-2}$	synonymous	0.960	0.255	$1.64 \times 10^{-4}$	LDLR	rs2569540	0.00	$1 \times 10^{-9}$	35589863
5	CBLC	rs80168591	19	44,781,370	G/a	1.4	0.87	splice	0.911	0.205	$8.54 \times 10^{-6}$	CBLC	rs899087	0.00	$5.1 \times 10^{-5}$	22832961
6	BCAM	rs28399637	19	44,820,881	G/a	31.5	$5 \times 10^{-8}$	intron	0.382	0.076	$4.71 \times 10^{-7}$	BCAM	rs2927480	0.21	$5.0 \times 10^{-49}$	21460841
7	PVRL2	rs283813	19	44,885,917	T/a	6.7	$3 \times 10^{-22}$	intron	0.488	0.123	$7.16 \times 10^{-5}$	PVRL2	rs283813	1	$7.6 \times 10^{-28}$	33589840
8	TOMM40	rs112849259	19	44,894,050	C/t	2.6	0.73	missense	1.227	0.138	$4.59 \times 10^{-19}$					
9	TOMM40	rs741780	19	44,901,174	T/c	43.2	$7 \times 10^{-2}$	intron	−0.454	0.080	$1.14 \times 10^{-8}$	TOMM40	rs741780	1	$1.4 \times 10^{-8}$	23565137
10	APOE	rs440446	19	44,905,910	G/c	35.8	$2 \times 10^{-2}$	missense	−0.561	0.087	$1.22 \times 10^{-10}$	APOE	rs439401	0.57	$1.1 \times 10^{-78}$	21460841
11	APOE	rs429358	19	44,908,684	T/c	15.4	0.54	missense	1.371	0.077	$4.63 \times 10^{-71}$	APOE	rs429358	1	$2.7 \times 10^{-78}$	21390209
12	APOE	rs7412	19	44,908,822	C/t	8.0	0.66	missense	−0.856	0.200	$1.87 \times 10^{-5}$	APOE	rs7412	1	$5.5 \times 10^{-58}$	20885792
13	APOC4; APOC2	rs5167	19	44,945,208	T/g	35.1	0.10	missense	0.315	0.075	$2.92 \times 10^{-5}$	APOC4; APOC2	rs5167	1	$2.8 \times 10^{-9}$	21460840

<sup>1</sup> Multiple genes were assigned if the index SNP was within the region of overlapping genes. <sup>2</sup> One single nucleotide polymorphism (SNP) per gene locus was retained. **Chr** = chromosome; **Ref/Alt** = reference/alternative allele (majuscule/minuscule letters stay for major/minor allele); **MAF obs** = minor allele frequency observer in our dataset. **Beta/SE/P** indicates effect size/standard error/*p*-value of the association of the alternative (minor) allele vs. the reference (major) allele. **SNP<sub>GRASP</sub>** and **Gene<sub>GRASP</sub>** denote SNP and related gene, for which minimum *p*-value has been reported in either the GRASP [32] or GWAS [33] catalog, where **P<sub>GRASP</sub>** is the reported *p*-value of the association with AD risk and **PMID** is the PubMed index of the respective paper. **r<sup>2</sup>** represents LD between previously reported SNP and SNP reported here. Empty cells: no index SNPs from the corresponding genetic locus were previously reported neither in the GRASP nor in GWAS catalog for their association with AD risk.



**Figure 1.** Flowchart of the analyses of this exome-wide association study (EWAS).

Outside of the *APOE* gene cluster, three SNPs demonstrated five GW significant associations with BMI (rs28688376), LDL-C, and TC (rs11244035, rs28997580) (Tables 4 and S2).

**Table 4.** Univariate and pair-wise pleiotropic associations of SNP, which includes Alzheimer’s disease and one of 16 considered phenotypes: 4 genetic loci outside the *APOE* gene region.

N	Gene(s) <sup>1</sup>	SNP <sup>2</sup>	Chr	Location, Base Pairs GRCh38	Ref /Alt	MAF obs (%)	PH	Beta	SE	P	P <sub>F</sub>
1	CDK11B (MMP23B)	rs28688376	1	1,637,577	T/c	26.1	BMI	−0.096	0.017	$3.36 \times 10^{-8}$	$1.42 \times 10^{-10}$
2	CDK11B (MMP23B)	rs28688376	1	1,637,577	T/c	26.1	Weight	−0.232	0.052	$7.01 \times 10^{-6}$	$2.37 \times 10^{-8}$
3	OBP2B	rs11244035	9	133,205,932	C/t	10.2	LDL-C	1.206	0.187	$1.09 \times 10^{-10}$	$1.13 \times 10^{-12}$
4	OBP2B	rs11244035	9	133,205,932	C/t	10.2	TC	1.519	0.243	$3.84 \times 10^{-10}$	$3.81 \times 10^{-12}$
5	TPM1	rs4775613	15	63,056,897	A/g	42.8	TC	−0.436	0.146	$2.81 \times 10^{-3}$	$9.91 \times 10^{-7}$
6	TPM1	rs4775613	15	63,056,897	A/g	42.8	HDL-C	−0.245	0.048	$2.43 \times 10^{-7}$	$1.31 \times 10^{-10}$
7	TPM1	rs4775613	15	63,056,897	A/g	42.8	SBP	0.156	0.062	$1.18 \times 10^{-2}$	$3.84 \times 10^{-6}$
8	TPM1	rs4775613	15	63,056,897	A/g	42.8	HF	−0.052	0.026	$4.16 \times 10^{-2}$	$1.24 \times 10^{-5}$
9	SMARCA4	rs28997580	19	11,013,062	C/t	0.9	LDL-C	−5.450	0.591	$2.81 \times 10^{-20}$	$<5 \times 10^{-8}$
10	SMARCA4	rs28997580	19	11,013,062	C/t	0.9	TC	−6.141	0.767	$1.20 \times 10^{-15}$	$<5 \times 10^{-8}$

<sup>1</sup> Multiple genes were assigned if the index SNP was within the region of overlapping genes. <sup>2</sup> One single nucleotide polymorphism (SNP) per gene locus was retained. Chr = chromosome; Ref/Alt = reference/alternative allele (majuscule/minuscule letters stay for major/minor allele); MAF obs = minor allele frequency observer in our dataset. Beta/SE/P indicates effect size/standard error/p-value of the association of the alternative (minor) allele vs. the reference (major) allele. P<sub>F</sub> is the p-value of the pair-wise pleiotropic analysis obtained by Fisher’s method.

In total, 29 GW significant univariate associations were identified.



### 3.2. Pleiotropic AD-Centric Pair-Wise Associations

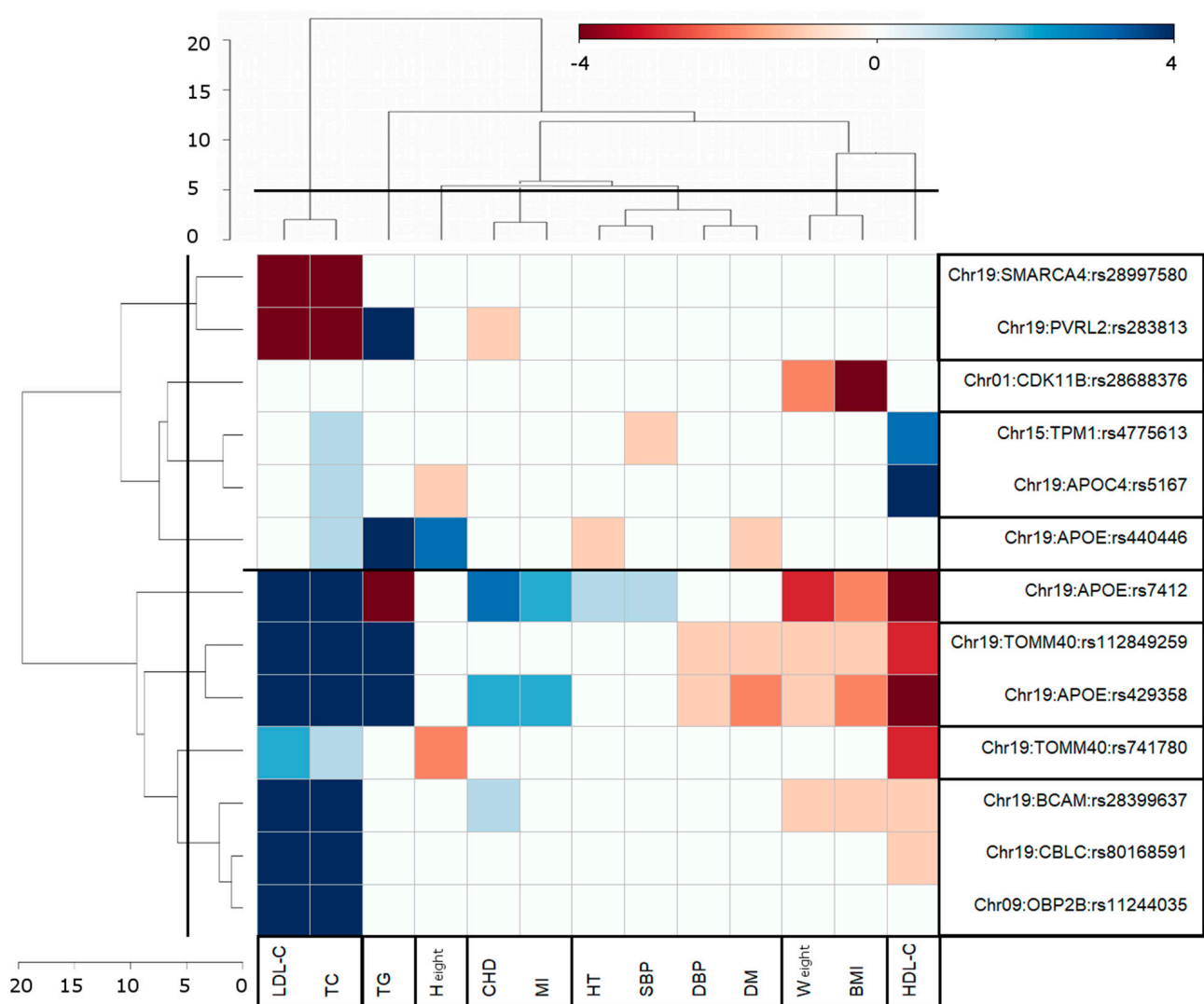
We identified 51 GW significant pleiotropic associations (see Section 2) (Tables S2 and S3A). They included 20 associations in the *APOE* gene cluster, which were not driven by the GW significance of the associations with AD. Four SNPs in the *APOE* and *TOMM40* genes, which were associated with AD at the GW level, showed 19 GW significant univariate associations with non-AD phenotypes, totaling 19 GW significant pleiotropic associations. These SNPs were associated with BMI, weight, and DBP (rs112849259 and rs429358); height and TC (rs741780 and rs440446); HDL-C (rs112849259 and rs741780); LDL-C (rs741780); DM (rs112849259, rs440446, and rs429358); HT (rs440446); and CHD and MI (rs429358). Also, the remaining SNPs from the *APOE* gene cluster showed five additional significant pleiotropic associations with AD and BMI, weight, CHD, and MI (rs7412); HDL-C (rs28399637), while these SNPs did not attain GW significance in the univariate analysis. In total, 44 [=20 + 19 + 5] GW significant pleiotropic associations were found for the SNPs mapped to the *APOE* gene cluster. Outside of the *APOE* cluster, we identified seven GW significant pair-wise pleiotropic associations. These included five SNP–trait pairs with GW significant univariate effects (Tables 4 and S2) and two additional pleiotropic associations, including BMI (rs28688376) and HDL-C (rs4775613).

We also identified 11 pleiotropic associations at the suggestive-effect level  $5 \times 10^{-8} < p < 5 \times 10^{-6}$ . They included associations of rs4775613 with SBP and TC; rs80168591 with HDL-C and stroke; rs28399637 with BMI, weight, and CHD; rs7412 with SBP and HT; as well as rs5167 with height and TC. Thus, in total, we found 62 (=51 + 11) pair-wise pleiotropic associations for 13 SNPs and 14 traits at  $p < 5 \times 10^{-6}$  (Tables S2 and S3A). None of the pleiotropic associations attained even a suggestive-effect level for BG and HF, and only one for stroke.

### 3.3. Clusters of Pleiotropic Associations

For cluster analysis (see Section 2.10), we used 61 pleiotropic associations, excluding unique association with stroke (Table S3A). Matrix  $A_{ij}$  for this analysis included information on 13 SNPs and 13 phenotypes (Table S3B). Hierarchical cluster analysis using the Euclidian measure and Ward's method provided two-dimensional dendrogram in the SNP–phenotype domain (Figure 2). Phenotype-specific one-dimensional dendrogram exemplified seven clusters for 13 phenotypes, i.e., (1) (LDL-C, TC), (2) TG, (3) (CHD, MI), (4) height, (5) (HT, SBP, DBP, DM), (6) HDL-C, and (7) (weight, BMI) (Figure 2). Pearson's correlation approach (Figure S2) provided similar results, except that a pair of (HT, SBP) phenotypes was clustered with (CHD, MI) rather than with (DBP, DM). This discrepancy is likely due to diminished power to identify clusters using different distance measures, given the moderate  $p$ -values from the univariate associations with these six phenotypes (Table S2). Both approaches of cluster analysis aggregated SNPs into the same eight clusters (Figures 2 and S2).

Figure 2 shows that the largest group of GW significant univariate (8 for LDL-C and 8 for TC) and pleiotropic (9 for LDL-C and 10 for TC) associations was observed for the (LDL-C, TC) cluster. The second largest group was observed for TG (five univariate and five pleiotropic GW significant associations), followed by HDL-C (three univariate and seven pleiotropic GW significant associations) and BMI (one univariate and four pleiotropic GW significant associations) + weight (four pleiotropic GW significant associations) clusters. Two distinct clusters emerged from the analysis, namely (LDL-C, TC) and (BMI, weight), which were composed of traits with the largest Pearson correlation coefficients of  $r = 0.952$  and  $r = 0.833$ , respectively, evaluated using individual-level data (Figure S1 and Table S1). This arrangement follows the anticipated clustering pattern where correlated traits are grouped together. However, in contrast, height and weight, despite having a substantial significant correlation of  $r = 0.538$ , were placed in separate phenotypic clusters.

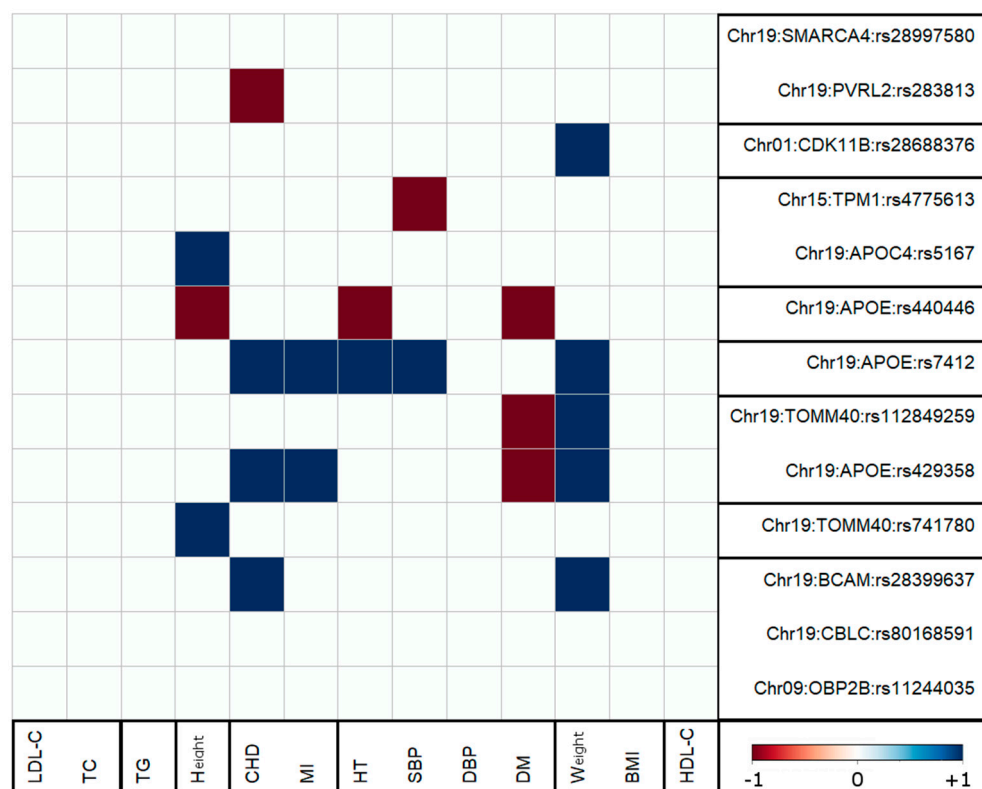


**Figure 2.** Eight clusters of SNPs and seven clusters of phenotypes based on significant pleiotropic pair-wise associations with AD and each of the other 13 age-related traits. Blue (red) color and its shades shows SNP–phenotype associations with the same (opposite) effect direction compared to the respective SNP–AD associations. Alleles, which demonstrated detrimental association with AD (positive betas), were considered effect alleles. Boxes show SNP and phenotypic clusters of pleiotropic associations, which were selected at a score (height) level of 5, as represented by axes with numbers on the horizontal (phenotypes) and vertical (SNPs) dendrograms. The  $x$ -axis shows the phenotypes, while  $y$ -axis refers to the SNPs and their respective genes. The horizontal solid line in the center shows the separation of the SNP clusters into two groups based on their significance and the effect directions of their associations with the first phenotypic cluster (LDL-C, TC). Hierarchical cluster algorithm with Euclidian measure and Ward’s method as implemented in R function `hclust` was used for this plot.

### 3.4. Antagonistic Genetic Heterogeneity

Figure 2 emphasizes the complex forms of genetic heterogeneity when the same alleles from different SNPs can have the same (e.g.,  $\epsilon 2$ -encoding rs7412 SNP) or opposite (e.g., minor allele of rs28997580) directions of the associations with AD and other phenotypes in a pair, e.g., (AD, LDL-C). This pattern, i.e., two sets of SNPs, was observed for each of the eight phenotypes from six phenotypic clusters: LDL-C + TC, TG, HDL-C, height (H), HT + SBP, and CHD.

Figure 3 provides further insight into the phenomenon of antagonistic genetic heterogeneity. It includes the results of the significant pleiotropic associations for the traits that demonstrated significant correlations with AD. Associations with lipid traits, BMI, and DBP were excluded because of their nonsignificant correlations with AD (Table 2). Each cell in Figure 3 displays the sign of a product of effect sizes for SNP–AD and SNP–trait associations and the Pearson correlation coefficient (calculated by using individual-level data) of AD and the corresponding trait. Red (blue) cells, which correspond to the negative (positive) sign of the product mentioned just above, represent pleiotropic associations with antagonistic (non-antagonistic) genetic effects. From this figure, it follows that 33% (7 out of 21) of the pleiotropic associations in 3 of 4 phenotypic clusters (75%) demonstrated antagonistic genetic heterogeneity, i.e., alleles, which demonstrated detrimental association with AD, also demonstrated the opposite sign direction of their association with the considered trait compared to the correlation of that trait with AD. Two traits demonstrated a more homogeneous pattern. Significant associations with DM demonstrated antagonistic pleiotropic effects only, while all significant pleiotropic associations, which included weight, were non-antagonistic ones (Figure 3).



**Figure 3.** Antagonistic genetic heterogeneity of significant pleiotropic pair-wise associations with AD and each of the other 13 age-related traits. Blue (red) color shows non-antagonistic (antagonistic) genetic heterogeneity. The results are presented for phenotypes that exhibit significant correlations with AD. Therefore, associations involving lipids, BMI, and DBP were excluded from this figure (see Table 2). Refer to Figure 2 caption for other notations.

#### 4. Discussion

In this EWAS, we performed AD-centric pair-wise pleiotropic and cluster analyses of the associations of SNPs with AD and 16 phenotypes of cardiovascular and AD risk factors.

##### 4.1. AD-Centric Pair-Wise Pleiotropic Associations

AD-centric pleiotropic analysis considered the associations of SNPs with AD and one of the 16 phenotypes. This analysis identified 13 SNPs with genome-wide significant

pleiotropic associations (Figure 2). Nine SNPs were mapped to the *APOE* gene cluster on chromosome 19 (including *APOE*, *TOMM40*, *APOC2/APOC4*, *PVRL2 (NECTIN2)*, *BCAM*, and *CBCL* genes), which is a well-known genetic risk factor for AD. Four other SNPs were from *CDK11* (rs28688376), *OBP2B* (rs11244035), *TPM1* (rs4775613), and *SMARCA4* (rs28997580) gene loci. These SNPs were associated with AD at  $p < 5 \times 10^{-4}$  and attained genome-wide significance ( $p < 5 \times 10^{-8}$ ) in the pleiotropic analysis.

The *CDK11B* (cyclin-dependent kinase 11B) gene plays a role in *cell apoptosis*. Recent studies have demonstrated that *CDK11* showed an altered expression in AD vulnerable neurons, which may be related to APP signaling processes [34,35]. A predominantly increased expression of *CDK11* was observed in the cytoplasm of neuronal cells in AD cases, while it was expressed specifically in the nuclei of post-mitotic neurons in most controls [34]. *CDK11* is regulated by checkpoint kinase 2 (CHK2), which phosphorylates tau at an AD-related site, enhancing tau toxicity [36,37], which demonstrates an additional link of the *CDK11B* gene to AD. Altered expression of *CDK11B* can be a cause of neuronal apoptosis and, consequently, of decreased brain weight through this process. This suggestion is consistent with our results that the minor allele of rs28688376 from the *CDK11B* gene was associated with an increased risk of AD and decreased levels of weight and BMI.

*TPM1*, *TPM3*, and *TPM4* genes encode tropomyosin isoforms in neuronal cells. *TPM1* products were found in the presynaptic compartment of the central nervous system (CNS) neurons [38]. Tropomyosins (Tpm) actin-binding proteins *stabilize the actin filaments*, which play a key role in the synaptic function of the CNS and mediate processes of memory and learning [39]. Microtubules, neurofilaments, and microfilaments (actin filaments) form the cytoskeleton of neurons. *Dysregulation* of tropomyosin and the actin *cytoskeleton can induce synapse loss*, which takes place in the early stages of AD pathology. Tropomyosin participation in the neurofibrillary pathology of AD was immunochemically demonstrated [40]. A significant increase in tropomyosin-1 abundance in the platelets of AD female patients was recently observed [41], in addition to a proteomics study that showed *TPM1* gene products increasing in the white matter of AD patients when compared to controls [42]. Recently, it has been demonstrated that *TPM1* plays a key role in cardio-genesis and cardiovascular disorders [43], providing a link to the associations of SNPs from this gene with SBP observed in this study.

The protein encoded by the *SMARCA4* gene is a member of the SWI/SNF family of proteins, which regulate gene activity by a chromatin remodeling and are involved in repairing damaged DNA, replicating DNA, and controlling the growth, division, and differentiation of cells. *SMARCA4* is in the same gene region as the *LDLR* gene. *LDLR* is a key regulator of cholesterol metabolism and encodes the protein involved in receptor-mediated endocytosis of low-density lipoprotein cholesterol [44]. The genome-wide-significant association of rs2569540 from the *LDLR* gene with AD was demonstrated in a recent GWAS study [45].

The *OBP2B* gene plays a role in chemosensory behavior and the perception of smell. Impaired sense of smell, or olfactory dysfunction, is often seen as an early indicator of AD pathology in the brain. A recent study indicated that individuals with olfactory dysfunction also exhibit changes in the blood levels of LDL-C and TC [46]. The current study aligns with these findings and adds additional insight, suggesting how blood cholesterol levels and AD-related changes could be intertwined. Specifically, our results suggest that olfactory dysfunction, which is considered as an early manifestation of AD, can be related to altered levels of LDL-C and TC in the blood. This link highlights the need for further investigation to determine whether AD might play a causal role in altering lipid metabolism.

Our findings suggest the involvement of other mechanisms contributing to the disease pathogenesis in addition to the amyloid-beta mechanism. Specifically, they highlight actin filaments, chromatin remodeling, neuronal apoptosis, and lipid abnormalities as potential targets for the development of new AD drugs.

#### 4.2. Clustering of Genome-Wide Significant Pair-Wise AD-Centric Pleiotropic Associations

In previous studies [11–17], the clustering of neurodegenerative conditions, including AD, was considered based on correlations among phenotypes. Our cluster analysis added a pleiotropic genetic component and, therefore, it can help in gaining further insights into the biological mechanisms associated with the pleiotropic effects. Such biological mechanisms may or may not be those which drive correlations among phenotypes. In this study, phenotypes belong to the same cluster if the associations of the SNPs with these phenotypes and AD demonstrated similarity defined by the same effect directions and the level of statistical significance (see Section 2.10). Thus, clustering is considered in a two-dimensional SNP/gene and phenotype domain. Similarity within phenotypic clusters suggests underlying biological mechanisms, which may contribute to AD. Our cluster analysis identified eight clusters of SNPs and seven clusters of phenotypes, which demonstrated similar SNP–phenotype associations (Figure 2). Cluster analysis was harmonized by considering alleles predisposed to AD as effect alleles.

We found that lipids defined three clusters: one cluster was defined by LDL-C and TC, while TG and HDL-C were in two separate clusters. This finding suggests at least partial independence of the biological mechanisms underlying the pleiotropic associations of each of these lipid traits and AD. This result corroborates previous studies [25,47], which reported partially independent mediation of the genetic associations with AD through LDL-C, TG, and HDL-C for SNPs inside and outside of the *APOE* gene region.

A CHD and MI cluster was defined by significant associations of SNPs from the *APOE* gene locus (Figure 2 and Table S2). Interestingly, while most alleles with elevated risks for CHD and MI (rs7412\_C:  $\beta_{\text{CHD}} = 0.12$ ,  $p = 8.25 \times 10^{-8}$ ; and rs423958\_c:  $\beta_{\text{CHD}} = 0.06$ ,  $p = 2.90 \times 10^{-4}$ ) were associated with higher risk of AD (rs7412\_C:  $\beta_{\text{AD}} = 0.86$ ,  $p = 1.87 \times 10^{-5}$ ; and rs423958\_c:  $\beta_{\text{AD}} = 1.37$ ,  $p = 4.63 \times 10^{-71}$ ), rs283813 (*PVRL2* gene) demonstrated the opposite relationship, i.e., rs283813\_a was favorably associated with CHD ( $\beta_{\text{CHD}} = -0.05$ ,  $p = 3.01 \times 10^{-2}$ ) but adversely with AD ( $\beta_{\text{AD}} = 0.49$ ,  $p = 7.16 \times 10^{-5}$ ). This finding demonstrates that alleles associated with increased risks of CHD can be associated with either a higher or lower risk of AD. This is in line with antagonistic pleiotropic associations identified for AD and other cardiovascular factors, such as DM, that likely reflects contributions of different biological processes [27,48,49].

Four traits (HT, SBP, DBP, DM) defined a cluster with a complex pattern of associations. As they are risk factors for CHD and MI, they are typically considered together with these diseases. Meanwhile, in this study, these phenotypes belong to two different clusters, as there is more similarity within the clusters than among the clusters. All six phenotypes and height would belong to the same cluster if the cut-off of the cluster height level was increased to 7 (Figure 2, see the vertical axis with the numbers in the top left corner). Interestingly, SNPs encoding *APOE*  $\epsilon 2$  and  $\epsilon 4$  alleles were associated with two different subclusters, (HT, SBP) and (DBP, DM), respectively. This finding emphasizes that the pleiotropic associations of these *APOE* alleles have different etiologies.

Height and (BMI, weight) defined two non-overlapping clusters, indicating a role of independent mechanisms contributing to height and BMI or weight and AD.

Three phenotypes (BG, HF, and stroke) did not define any phenotypic cluster because of the lack of significant genetic associations that could be used for clustering.

#### 4.3. Antagonistic Genetic Heterogeneity Was Observed for AD-Centric Pleiotropic Associations with Five Traits

Antagonistic genetic heterogeneity is characterized by misalignment of the signs of the product of the effects of genetic associations with different phenotypes and the correlations among these phenotypes [22–26]. This heterogeneity in our study was seen in the associations of seven SNPs with five traits in three phenotypic clusters (Figure 3). For each cluster of phenotypes (apart BMI, weight and DBP, DM), there were two sets of SNPs, which were characterized by antagonistic and non-antagonistic genetic effects in the associations with AD and the second phenotype in a pair (Figure 3). For example,

rs283813 showed antagonistic genetic heterogeneity because it was adversely associated with AD and favorably with CHD, despite a significant positive correlation between AD and CHD (Table 2). The rs429358 did not show antagonistic heterogeneity with AD and weight because opposite-direction associations of this SNP with these phenotypes were aligned with a negative correlation between AD and weight.

Antagonistic genetic heterogeneity can also be viewed as a source of SNP/gene-phenotype clustering. For example, antagonistic genetic heterogeneity differentiates a cluster of height from that of BMI and weight because the height cluster includes antagonistic heterogeneity whereas the latter does not.

Antagonistic genetic heterogeneity plays a substantial role in the genetic associations with AD and its risk factors. This heterogeneity indicates that an effect allele from an SNP can confer the risk of one trait while being favorably associated with the other traits. Such a relationship likely indicates the role of different biological processes associated with the genes and pathways mapped to such SNPs on AD and its risk factors. For example, for DM, all significant pleiotropic associations were antagonistic. The alleles of SNPs with a significant adverse effect on DM were favorably associated with AD. This finding corroborates previous results [27,48,49] and highlights the complex contribution of genetic components to these diseases.

#### 4.4. Protective Effect of Higher BMI Level against AD Is Related to High Weight

Previous studies demonstrated the protective effect of a high BMI level in late life against AD [20,21]. BMI is defined by two phenotypes, weight and height. Our cluster analysis showed that weight and height belong to different clusters. Moreover, SNPs that demonstrated genome-wide significant pleiotropic associations with BMI and weight were not significantly associated with height and vice versa. This finding suggests that protective effects of a higher BMI level against AD are related to higher weight rather than smaller height. While an increased BMI resulting from weight gain is associated with a higher risk of cardiovascular diseases (such as CHD, MI, and hypertension) and diabetes in midlife, a decreased BMI resulting from weight loss in later life can increase the risk of AD. This risk is attributed to the deteriorated functions of an aging organism, which include impaired metabolism, reduced cholesterol levels, decreased muscle mass, frailty, and sarcopenia, among other factors.

## 5. Conclusions

This study reports four novel genetic loci showing pleiotropic associations with AD and at least with one of 13 cardiovascular and AD risk factors. Genes harboring these loci are involved in cell apoptosis (*CDK11*), the stabilization of actin filaments, i.e., the cytoskeleton of neurons, (*TPM1*), the regulation of chromatin remodeling (*SMARCA4*), and chemosensory behavior and sensory perception of smell (*OBP2B*). No pleiotropic associations were identified for BG, HF, or stroke. Leveraging information on univariate and pleiotropic genetic associations, we found seven clusters in the domain of 13 SNPs and 13 phenotypes. Nine of thirteen pleiotropic SNPs were mapped to the *APOE* gene cluster (Figure 2).

AD-centric pleiotropic analysis confirmed partially independent mechanisms of pleiotropic associations with AD and lipid traits; although LDL-C and TC defined one phenotypic cluster, TG and HDL-C formed independent clusters because of the different patterns of the genetic associations.

Our analysis identified that 61% (9 of 13) of the considered traits demonstrated antagonistic pleiotropic AD-centric associations. Our cluster analysis found that weight, which was clustered with BMI, but not with height, defined pleiotropy of BMI and AD. SNPs identified by significant pleiotropic associations with AD and DM were from the *APOE* gene cluster, which corroborates previous findings.

Our cluster analysis in a two-dimensional SNP/gene-phenotype domain highlighted a more complex role of genetic and non-genetic factors in AD pathogenesis than the cluster analyses based on phenotype correlations did. Indeed, assuming that the clustering of

phenotypes is related to their correlations, it is expected that correlated phenotypes should be in the same cluster. For LDL-C and TC (Pearson's correlation coefficient,  $r = 0.952$ ), BMI and weight ( $r = 0.833$ ), and CHD and MI ( $r = 0.581$ ), this was the case, as each pair of these phenotypes was in its own cluster. This was not the case, however, for the other correlated phenotypes, such as, for example, height and weight ( $r = 0.538$ ), HDL-C and weight ( $r = -0.462$ ), and HDL-C and TG ( $r = -0.439$ ) (see Table S1 and Figure S1) because these phenotypes were in different clusters. Accordingly, our cluster analysis approach was capable of capturing the biological mechanisms relevant to pleiotropy underlying AD and other age-related phenotypes that may be not readily available from analyses relying on information about the correlations of phenotypes. Defining more homogeneous patterns of phenotypes and genetic factors can aid in the identification of more homogeneous groups of individuals who are at varying risks of AD. Subsequent research can facilitate the identification of genetic and non-genetic components, contributing to the pleiotropic effects associated with AD. Such insights can be valuable for developing comprehensive treatment approaches that combine drug therapy and lifestyle interventions.

The proposed approach can be applied to published summary statistics of genetic associations with different traits. Such an extension can help identify phenotypes that share genetic components and, therefore, overlapping biological mechanisms between different complex traits.

This study has some limitations. First, the analyses were performed for the entire sample but not separately for groups that are at different risk levels of AD, such as, for instance, within each sex group and/or at different ages. Second, we used a threshold  $p$ -value =  $5 \times 10^{-4}$  for associations with AD, which substantially limited the number of significant pleiotropic associations/SNPs. Third, a small number of SNPs (because of the second limitation) were available for hierarchical cluster analysis, which limited the precision of identifying phenotype clustering.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14101834/s1>, Table S1: Pearson's correlation coefficient [%] between 17 considered traits including cardiovascular and Alzheimer's disease risk factors in the UK Biobank; Table S2: Univariate and pleiotropic associations of 13 SNPs in the APOE gene region and other 4 gene loci, which demonstrated genome-wide significance of pair-wise pleiotropic associations with Alzheimer's disease and at least one of 16 traits; Table S3: (A) Patterns/clusters of pleiotropic associations of 13 SNPs in the APOE gene region and the other 4 gene loci, which demonstrated genome-wide significance of pair-wise pleiotropic associations with Alzheimer's disease and at least one of 16 traits. (B) Matrix A that was used in the cluster analysis; Figure S1: Pearson's correlation coefficients between 17 considered traits, which represent cardiovascular and Alzheimer's disease risk factors in the UK Biobank; Figure S2: Clusters of 13 traits and 13 SNPs based on Pearson's correlation coefficients as a distance measure as implemented in R function `phetmap` from the R package `ecodist`.

**Author Contributions:** Investigation, Data Curation, Validation, Writing—Original Draft Preparation, Writing—Review and Editing: Y.L., E.L., I.C. and A.M.K.; Bioinformatic and Biological Analysis: E.L. and I.C.; Conceptualization, Methodology, and Visualization: Y.L. and A.M.K.; Software and Formal Analysis: Y.L.; Funding Acquisition, Supervision, and Project Administration: A.M.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Grants R01AG065477 and R01AG070488 from the National Institute on Aging. The funding source did not participate in the writing of the manuscript or the decision to submit the manuscript for publication.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Duke University (Pro00105247-INIT-1.0 approved on 26 June 2020; Pro00105346-INIT-1.0 approved on 15 April 2020).

**Informed Consent Statement:** Informed consent was obtained from all subjects by primary UKB investigators who gathered data.

**Data Availability Statement:** This manuscript was prepared using limited-access phenotypic and genetic datasets of the UK Biobank, which are available through the following link: <http://www.ukbiobank.ac.uk/>, accessed on 15 March 2021.

**Acknowledgments:** This research was conducted using data from UK Biobank, a major biomedical database (<http://www.ukbiobank.ac.uk/>, accessed on 15 March 2021). See also the Supplementary Materials.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. Gaugler, J.; James, B.; Johnson, T.; Reimer, J.; Solis, M.; Weuve, J. 2022 Alzheimer's disease facts and figures. *Alzheimers Dement.* **2022**, *18*, 700–789. [[CrossRef](#)]
2. Malik, R.; Kalra, S.; Bhatia, S.; Harrasi, A.A.; Singh, G.; Mohan, S.; Makeen, H.A.; Albratty, M.; Meraya, A.; Bahar, B.; et al. Overview of therapeutic targets in management of dementia. *Biomed. Pharmacother.* **2022**, *152*, 113168. [[CrossRef](#)] [[PubMed](#)]
3. Cummings, J. Anti-Amyloid Monoclonal Antibodies are Transformative Treatments that Redefine Alzheimer's Disease Therapeutics. *Drugs* **2023**, *83*, 569–576. [[CrossRef](#)]
4. Ferrer, I. Hypothesis review: Alzheimer's overtone guidelines. *Brain Pathol.* **2023**, *33*, e13122. [[CrossRef](#)]
5. Pons, V.; Rivest, S. Targeting Systemic Innate Immune Cells as a Therapeutic Avenue for Alzheimer Disease. *Pharmacol. Rev.* **2022**, *74*, 1–17. [[CrossRef](#)]
6. Tang, Q.; Wang, F.; Yang, J.; Peng, H.; Li, Y.; Li, B.; Wang, S. Revealing a Novel Landscape of the Association Between Blood Lipid Levels and Alzheimer's Disease: A Meta-Analysis of a Case-Control Study. *Front. Aging Neurosci.* **2019**, *11*, 370. [[CrossRef](#)] [[PubMed](#)]
7. Anstey, K.J.; Ashby-Mitchell, K.; Peters, R. Updating the Evidence on the Association between Serum Cholesterol and Risk of Late-Life Dementia: Review and Meta-Analysis. *J. Alzheimer's Dis.* **2017**, *56*, 215–228. [[CrossRef](#)] [[PubMed](#)]
8. Zhang, X.; Tong, T.; Chang, A.; Ang, T.F.A.; Tao, Q.; Auerbach, S.; Devine, S.; Qiu, W.Q.; Mez, J.; Massaro, J.; et al. Midlife lipid and glucose levels are associated with Alzheimer's disease. *Alzheimer's Dement.* **2023**, *19*, 181–193. [[CrossRef](#)]
9. Bone, W.P.; Siewert, K.M.; Jha, A.; Klarin, D.; Damrauer, S.M.; Program, V.A.M.V.; Chang, K.M.; Tsao, P.S.; Assimes, T.L.; Ritchie, M.D.; et al. Multi-trait association studies discover pleiotropic loci between Alzheimer's disease and cardiometabolic traits. *Alzheimer's Res. Ther.* **2021**, *13*, 34. [[CrossRef](#)]
10. Broce, I.J.; Tan, C.H.; Fan, C.C.; Jansen, I.; Savage, J.E.; Witoelar, A.; Wen, N.; Hess, C.P.; Dillon, W.P.; Glastonbury, C.M.; et al. Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. *Acta Neuropathol.* **2019**, *137*, 209–226. [[CrossRef](#)]
11. Luchsinger, J.A.; Reitz, C.; Honig, L.S.; Tang, M.X.; Shea, S.; Mayeux, R. Aggregation of vascular risk factors and risk of incident Alzheimer disease. *Neurology* **2005**, *65*, 545–551. [[CrossRef](#)]
12. Kivipelto, M.; Ngandu, T.; Fratiglioni, L.; Viitanen, M.; Kareholt, I.; Winblad, B.; Helkala, E.L.; Tuomilehto, J.; Soininen, H.; Nissinen, A. Obesity and vascular risk factors at midlife and the risk of dementia and Alzheimer disease. *Arch. Neurol.* **2005**, *62*, 1556–1560. [[CrossRef](#)]
13. Xiang, Q.; Andersen, S.L.; Sweigart, B.; Gunn, S.; Nygaard, M.; Perls, T.T.; Sebastiani, P. Signatures of Neuropsychological Test Results in the Long Life Family Study: A Cluster Analysis. *J. Alzheimer's Dis.* **2023**, *93*, 1457–1469. [[CrossRef](#)]
14. Ramoni, M.F.; Sebastiani, P.; Kohane, I.S. Cluster analysis of gene expression dynamics. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9121–9126. [[CrossRef](#)] [[PubMed](#)]
15. Bondi, M.W.; Edmonds, E.C.; Jak, A.J.; Clark, L.R.; Delano-Wood, L.; McDonald, C.R.; Nation, D.A.; Libon, D.J.; Au, R.; Galasko, D.; et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J. Alzheimer's Dis.* **2014**, *42*, 275–289. [[CrossRef](#)]
16. Edmonds, E.C.; McDonald, C.R.; Marshall, A.; Thomas, K.R.; Eppig, J.; Weigand, A.J.; Delano-Wood, L.; Galasko, D.R.; Salmon, D.P.; Bondi, M.W.; et al. Early versus late MCI: Improved MCI staging using a neuropsychological approach. *Alzheimer's Dement.* **2019**, *15*, 699–708. [[CrossRef](#)]
17. Jak, A.J.; Preis, S.R.; Beiser, A.S.; Seshadri, S.; Wolf, P.A.; Bondi, M.W.; Au, R. Neuropsychological Criteria for Mild Cognitive Impairment and Dementia Risk in the Framingham Heart Study. *J. Int. Neuropsychol. Soc.* **2016**, *22*, 937–943. [[CrossRef](#)] [[PubMed](#)]
18. Kulmiski, A.M. Genetic Association Studies of Age-Related Traits: New Perspectives. *Adv. Geriatr. Med. Res.* **2021**, *3*, e210003. [[CrossRef](#)] [[PubMed](#)]
19. Woodward, A.A.; Urbanowicz, R.J.; Naj, A.C.; Moore, J.H. Genetic heterogeneity: Challenges, impacts, and methods through an associative lens. *Genet. Epidemiol.* **2022**, *46*, 555–571. [[CrossRef](#)]
20. Pedditzi, E.; Peters, R.; Beckett, N. The risk of overweight/obesity in mid-life and late life for the development of dementia: A systematic review and meta-analysis of longitudinal studies. *Age Ageing* **2016**, *45*, 14–21. [[CrossRef](#)]
21. Tolppanen, A.M.; Ngandu, T.; Kareholt, I.; Laatikainen, T.; Rusanen, M.; Soininen, H.; Kivipelto, M. Midlife and late-life body mass index and late-life dementia: Results from a prospective population-based cohort. *J. Alzheimer's Dis.* **2014**, *38*, 201–209. [[CrossRef](#)]



22. Kulminski, A.M.; Loika, Y.; Huang, J.; Arbeev, K.G.; Bagley, O.; Ukraintseva, S.; Yashin, A.I.; Culminskaya, I. Pleiotropic Meta-Analysis of Age-Related Phenotypes Addressing Evolutionary Uncertainty in Their Molecular Mechanisms. *Front. Genet.* **2019**, *10*, 433. [[CrossRef](#)]
23. Kulminski, A.M.; Huang, J.; Loika, Y.; Arbeev, K.G.; Bagley, O.; Yashkin, A.; Duan, M.; Culminskaya, I. Strong impact of natural-selection-free heterogeneity in genetics of age-related phenotypes. *Aging* **2018**, *10*, 492–514. [[CrossRef](#)]
24. Kulminski, A.M.; Loika, Y.; Nazarian, A.; Culminskaya, I. Quantitative and Qualitative Role of Antagonistic Heterogeneity in Genetics of Blood Lipids. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2020**, *75*, 1811–1819. [[CrossRef](#)]
25. Loika, Y.; Feng, F.; Loiko, E.; Kulminski, A.M. Mediation of the APOE associations with Alzheimer’s and coronary heart diseases through body mass index and lipids. *GeroScience* **2022**, *44*, 1141–1156. [[CrossRef](#)] [[PubMed](#)]
26. Kulminski, A.M.; Loiko, E.; Loika, Y.; Culminskaya, I. Pleiotropic predisposition to Alzheimer’s disease and educational attainment: Insights from the summary statistics analysis. *GeroScience* **2022**, *44*, 265–280. [[CrossRef](#)] [[PubMed](#)]
27. Kulminski, A.M.; Feng, F.; Loiko, E.; Nazarian, A.; Loika, Y.; Culminskaya, I. Prevailing Antagonistic Risks in Pleiotropic Associations with Alzheimer’s Disease and Diabetes. *J. Alzheimer’s Dis.* **2023**, *94*, 1121–1132. [[CrossRef](#)] [[PubMed](#)]
28. Sudlow, C.; Gallacher, J.; Allen, N.; Beral, V.; Burton, P.; Danesh, J.; Downey, P.; Elliott, P.; Green, J.; Landray, M.; et al. UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **2015**, *12*, e1001779. [[CrossRef](#)]
29. Bycroft, C.; Freeman, C.; Petkova, D.; Band, G.; Elliott, L.T.; Sharp, K.; Motyer, A.; Vukcevic, D.; Delaneau, O.; O’Connell, J.; et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **2018**, *562*, 203–209. [[CrossRef](#)]
30. Szustakowski, J.D.; Balasubramanian, S.; Kvikstad, E.; Khalid, S.; Bronson, P.G.; Sasson, A.; Wong, E.; Liu, D.; Wade Davis, J.; Haefliger, C.; et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat. Genet.* **2021**, *53*, 942–948. [[CrossRef](#)]
31. Fisher, R.A.S. *Statistical Methods for Research Workers*, 14th ed.; Oliver and Boyd: Edinburgh, UK, 1970.
32. Leslie, R.; O’Donnell, C.J.; Johnson, A.D. GRASP: Analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* **2014**, *30*, i185–i194. [[CrossRef](#)]
33. Sollis, E.; Mosaku, A.; Abid, A.; Buniello, A.; Cerezo, M.; Gil, L.; Groza, T.; Gunes, O.; Hall, P.; Hayhurst, J.; et al. The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource. *Nucleic Acids Res.* **2023**, *51*, D977–D985. [[CrossRef](#)] [[PubMed](#)]
34. Bajic, V.P.; Su, B.; Lee, H.G.; Kudo, W.; Siedlak, S.L.; Zivkovic, L.; Spremo-Potparevic, B.; Djelic, N.; Milicevic, Z.; Singh, A.K.; et al. Mislocalization of CDK11/PITSLRE, a regulator of the G2/M phase of the cell cycle, in Alzheimer disease. *Cell. Mol. Biol. Lett.* **2011**, *16*, 359–372. [[CrossRef](#)] [[PubMed](#)]
35. Fredi, B.M.; De Labio, R.W.; Rasmussen, L.T.; Chagas, E.F.B.; Chen, E.S.; Turecki, G.; Smith, M.A.C.; Payao, S.L.M. CDK10, CDK11, FOXO1, and FOXO3 Gene Expression in Alzheimer’s Disease Encephalic Samples. *Cell. Mol. Neurobiol.* **2023**, *43*, 2953–2962. [[CrossRef](#)]
36. Choi, H.H.; Choi, H.K.; Jung, S.Y.; Hyle, J.; Kim, B.J.; Yoon, K.; Cho, E.J.; Youn, H.D.; Lahti, J.M.; Qin, J.; et al. CHK2 kinase promotes pre-mRNA splicing via phosphorylating CDK11(p110). *Oncogene* **2014**, *33*, 108–115. [[CrossRef](#)] [[PubMed](#)]
37. Mendoza, J.; Sekiya, M.; Taniguchi, T.; Iijima, K.M.; Wang, R.; Ando, K. Global analysis of phosphorylation of tau by the checkpoint kinases Chk1 and Chk2 in vitro. *J. Proteome Res.* **2013**, *12*, 2654–2665. [[CrossRef](#)]
38. Guven, K.; Gunning, P.; Fath, T. TPM3 and TPM4 gene products segregate to the postsynaptic region of central nervous system synapses. *Bioarchitecture* **2011**, *1*, 284–289. [[CrossRef](#)]
39. Pelucchi, S.; Stringhi, R.; Marcello, E. Dendritic Spines in Alzheimer’s Disease: How the Actin Cytoskeleton Contributes to Synaptic Failure. *Int. J. Mol. Sci.* **2020**, *21*, 908. [[CrossRef](#)]
40. Galloway, P.G.; Mulvihill, P.; Siedlak, S.; Mijares, M.; Kawai, M.; Padget, H.; Kim, R.; Perry, G. Immunochemical demonstration of tropomyosin in the neurofibrillary pathology of Alzheimer’s disease. *Am. J. Pathol.* **1990**, *137*, 291–300.
41. Reumiller, C.M.; Schmidt, G.J.; Dhrami, I.; Umlauf, E.; Rappold, E.; Zellner, M. Gender-related increase of tropomyosin-1 abundance in platelets of Alzheimer’s disease and mild cognitive impairment patients. *J. Proteom.* **2018**, *178*, 73–81. [[CrossRef](#)]
42. Castano, E.M.; Maarouf, C.L.; Wu, T.; Leal, M.C.; Whiteside, C.M.; Lue, L.F.; Kokjohn, T.A.; Sabbagh, M.N.; Beach, T.G.; Roher, A.E. Alzheimer disease periventricular white matter lesions exhibit specific proteomic profile alterations. *Neurochem. Int.* **2013**, *62*, 145–156. [[CrossRef](#)] [[PubMed](#)]
43. England, J.; Granados-Riveron, J.; Polo-Parada, L.; Kuriakose, D.; Moore, C.; Brook, J.D.; Rutland, C.S.; Setchfield, K.; Gell, C.; Ghosh, T.K.; et al. Tropomyosin 1: Multiple roles in the developing heart and in the formation of congenital heart defects. *J. Mol. Cell. Cardiol.* **2017**, *106*, 1–13. [[CrossRef](#)] [[PubMed](#)]
44. Brown, M.S.; Goldstein, J.L. Receptor-mediated endocytosis: Insights from the lipoprotein receptor system. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 3330–3337. [[CrossRef](#)]
45. Gouveia, C.; Gibbons, E.; Dehghani, N.; Eapen, J.; Guerreiro, R.; Bras, J. Genome-wide association of polygenic risk extremes for Alzheimer’s disease in the UK Biobank. *Sci. Rep.* **2022**, *12*, 8404. [[CrossRef](#)]
46. Gallo, S.; Byham-Gray, L.; Duffy, V.B.; Hoffman, H.J.; Hayes, J.E.; Rawal, S. Associations of olfactory dysfunction with anthropometric and cardiometabolic measures: Findings from the 2013–2014 national health and nutrition examination survey (NHANES). *Physiol. Behav.* **2020**, *215*, 112702. [[CrossRef](#)] [[PubMed](#)]

47. Loika, Y.; Loiko, E.; Feng, F.; Stallard, E.; Yashin, A.I.; Arbee, K.; Kuipers, A.L.; Feitosa, M.F.; Province, M.A.; Kulminski, A.M. Exogenous exposures shape genetic predisposition to lipids, Alzheimer's, and coronary heart disease in the MLXIPL gene locus. *Aging* **2023**, *15*, 3249–3272. [[CrossRef](#)]
48. Hao, K.; Di Narzo, A.F.; Ho, L.; Luo, W.; Li, S.; Chen, R.; Li, T.; Dubner, L.; Pasinetti, G.M. Shared genetic etiology underlying Alzheimer's disease and type 2 diabetes. *Mol. Asp. Med.* **2015**, *43–44*, 66–76. [[CrossRef](#)]
49. Hardy, J.; de Strooper, B.; Escott-Price, V. Diabetes and Alzheimer's disease: Shared genetic susceptibility? *Lancet Neurol.* **2022**, *21*, 962–964. [[CrossRef](#)]

**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.