

---

# ApoE4 dose effects on serum metabolites in Alzheimer's disease

a Data Science approach

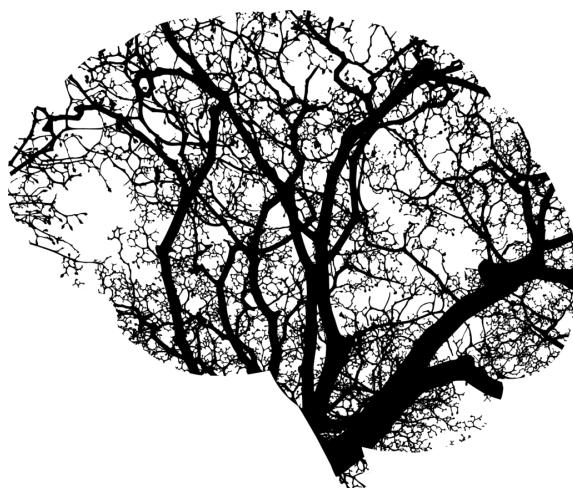
---

George Miliarakis

MSc Thesis  
Wageningen University & Research  
Wageningen, March 2024

*Supervisor*  
**C.F.W. Peeters**  
Mathematical & Statistical Methods (Biometris)  
Wageningen University & Research

*Supervisor*  
**Yannick Vermeiren**  
Nutritional Biology, Human Nutrition & Health  
Wageningen University & Research



## ABBREVIATIONS

- ABCA1:** ATP-binding cassette transporter A1.  
**AD:** Alzheimer's disease.  
**ApoE:** apolipoprotein E.  
**APP:** amyloid precursor protein.  
**ATP:** adenosine tri-phosphate.  
**AUC:** area under the ROC curve.
- BBB:** blood-brain barrier.  
**CSF:** cerebrospinal fluid.  
**CV:** cross-validation.
- DT:** decision tree.
- FAD:** familial Alzheimer's disease.
- GGM:** Gaussian graphical model.  
**GSK-3 $\beta$ :** glycogen synthase kinase-3 $\beta$ .
- HDL:** high density lipoprotein.
- IDL:** intermediate density lipoprotein.  
**IL:** interleukin.  
**ISF:** interstitial fluid.
- LDLR:** low-density lipoprotein receptor.  
**LOAD:** late-onset Alzheimer's disease.  
**LRP1:** low density lipoprotein receptor-related protein 1.
- MCI:** mild cognitive impairment.  
**ML:** maximum likelihood.  
**MLR:** multinomial logistic regression.
- NFkB:** nuclear factor kappa B.
- ROC:** receiver-operating characteristics.  
**ROS:** reactive oxygen species.
- SAD:** sporadic Alzheimer's disease.  
**SCD:** subjective cognitive decline.  
**SMOTE:** synthetic minority over-sampling technique.
- TLR4:** toll-like receptor 4.  
**TNF- $\alpha$ :** tumor necrosis factor  $\alpha$ .
- VCS:** version control system.  
**VLDL:** very low density lipoprotein.
- XGBoost:** eXtreme gradient boosting.

## CONTENTS

Abbreviations	i
1. Introduction	1
1.1. Alzheimer's disease	1
1.2. Human apolipoprotein-E gene	1
1.3. Apolipoprotein E	3
1.4. ApoE4-mediated metabolic changes in Alzheimer's disease	5
1.5. Research Questions	7
1.6. Approach and Overview	7
2. Methods	8
2.1. Subjects	8
2.2. Data management	8
2.3. Feature Engineering	8
2.4. Statistical Analysis	9
3. Results	14
3.1. ApoE4 dose effects on serum metabolite levels in AD	14
3.2. Classification of ApoE4 and AD status	17
3.3. Metabolite Covariance Network Analysis	20
4. Discussion	23
5. Conclusion	25
References	26
Appendix A. Multiclass ROC curves	32
Appendix B. R Session Information	34

## 1. INTRODUCTION

**1.1. Alzheimer's disease.** Alzheimer's disease (AD) is a complex, progressive neurodegenerative disorder and the most common form of dementia [1]. It was considered the 6th leading cause of death in the US in 2019, with an overall increase of 145% in mortality from 2000 to 2019 [2]. The impact AD has on patients, their caregivers, and healthcare systems is detrimental. Hence, a considerable amount of research has been performed in an effort to understand, prevent, impede or cure it. Nonetheless, important aspects of its systemic manifestations remain unknown.

Age is the main risk factor for AD, while several genetic and lifestyle risk factors, as well as biochemical pathways contribute to its development [1]. AD manifests in various histopathological phenotypes and presents a broad spectrum of clinical signs and symptoms [3, 4]. The AD continuum starts with subjective cognitive decline (SCD), followed by mild cognitive impairment (MCI) [5], and continues with progressive loss of global cognition, of which particularly memory, processing speed and executive functioning, spanning a total period of 10-15 years [6].

Sporadic or late-onset AD (SAD, LOAD; 95% of cases) is the most frequent phenotype, typically appearing after 65 years of age [7]. A rarer phenotype is early-onset familial AD (FAD), usually starting at ages 30–65 and passed in an autosomal dominant fashion [8]. Even though FAD mutations explain only a small percentage of AD cases, they have a great impact on AD research given their appealing genotype-phenotype links.

*Amyloid cascade hypothesis.* The amyloid cascade hypothesis has dominated the scientific discussion on the pathogenesis of AD. In historical terms, its impact was profound, as it helped distinguish and identify AD as a single disease that may be studied for treatment [9]. It suggests that chronic neuroinflammation promotes protein misfolding and accumulation in the brain, forming plaques (consisting of oligomerized amyloid  $\text{A}\beta_{42}$ ) and tangles (consisting of hyperphosphorylated tau protein) [4]. Nevertheless, it does not necessarily cover all AD cases. Clinical trials with  $\text{A}\beta$  antibodies as a potential disease-modifying treatment have so far failed to demonstrate any clinical improvement in patients, despite successful removal of plaques from AD brains [10, 11]. Evidence suggests that oxidative stress, metabolic abnormalities, atherosclerosis, cardiovascular effects, imbalances of intra-neuronal calcium and metal ion deposition might also contribute to the development of AD [10]. Kepp *et al.* propose a more complex and holistic view of AD pathology, by integrating (epi-)genetic, environmental, vascular, neuro-inflammatory and metabolic factors in disease-predicting models [10].

**1.2. Human apolipoprotein-E gene.** LOAD has been associated with genetic risk factors, such as with the gene that encodes for apolipoprotein E (ApoE) [12]. The structure and function of ApoE, as well as how the first defines the latter is described in Section 1.3.

*ApoE4 and Alzheimer's disease.* Humans present three variants of the ApoE gene:  $\varepsilon_2$ ,  $\varepsilon_3$  and  $\varepsilon_4$  [13, 14], resulting in six genotypes [Fig. 1]. In Caucasian populations, the most abundant allele is ApoE3 (rs7412 C/rs429358 T), with a frequency of 78%, and is considered neutral regarding the risk of AD [15]. ApoE4 (rs7412 C/rs429358 C) has a frequency of 14% and represents the strongest genetic risk factor for LOAD, with gene dose effects [16]. Conversely, ApoE2 (rs7412 T/rs429358 T) is found in 8% of Caucasian populations, and is associated with a reduced risk of LOAD [15]. The various ApoE alleles are strongly linked to the primary pathological features of LOAD, namely  $\text{A}\beta$  and phosphorylated tau [17]. While the association between ApoE alleles and LOAD risk or protection is observed across diverse ancestral backgrounds, the strength of this association varies [18, 19], see Section 1.2.2.

Carrying a single  $\varepsilon_4$  allele implies a risk of developing AD of about 20% [20]. A heterozygous ApoE4 allelic composition (e.g.,  $\varepsilon_4/\varepsilon_x$ ) is associated with approximately 3-4 times increased life-time risk of developing AD, while the homozygous composition ( $\varepsilon_4/\varepsilon_4$ ) 12 times increased [21]. Even though ApoE4 is the strongest known genetic risk factor for LOAD, it is neither necessary nor sufficient to cause AD, and it is certainly not the only genetic risk factor [22].

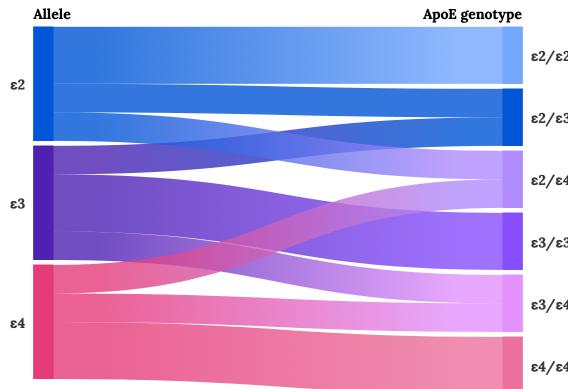


FIGURE 1. Sankey chart reflecting the distribution of ApoE alleles in the composition of the 6 ApoE genotypes.

*ApoE and sex synergy in Alzheimer's disease.* Notably, sex (60% females) and ApoE4 allelic composition (50% has at least one  $\epsilon_4$  allele) are the strongest genetic risk factors for LOAD [23]. In this regard, it is shown that the ApoE4 genotype has a larger impact on females, as they present greater impairment of mitochondrial energy production, compared to males [23, 24].

*ApoE and ancestry in Alzheimer's disease.* The majority of studies exploring the relationship between ApoE alleles and the genetics of LOAD have primarily focused on Northern European populations [14]. However, smaller studies involving diverse ancestral backgrounds show variations in ApoE4 allele [14]. While ApoE4 is present in 14% of Caucasian Americans, its prevalence increases to 40% among African Americans, 37% in Oceania, and 26% in Australia. Southern Asia and Europe exhibit ApoE4 allelic frequencies of less than 10%, compared to Northern Europe where it rises to 25% [18, 25–27].

The epidemiological impact of ApoE alleles also differs among populations. In Korea, Japan, and Japanese-American communities, ApoE4 confers a higher risk of LOAD compared to Caucasians [19]. Conversely, for Native Americans, Hispanic Americans, African Americans, and those of African descent, ApoE4 is associated with a lower risk of LOAD than in Caucasian-American populations [19, 28–31]. A recent study in a Chinese population found that ApoE3 is more protective than ApoE2 [32]. Some of these population-specific effects are attributed to the ApoE haplotype [28, 30]. A recent discovery of a novel locus (19q13.31) could contribute to attenuating ApoE4-mediated AD risk in African Americans [33].

*Evolution of ApoE over time.* Interestingly, humans are the only species exhibiting polymorphism in the ApoE gene [24]. All other animal species have one ApoE variant, which resembles the human ApoE3 allele [34]. ApoE4 is the oldest human allele, followed by ApoE3 and ApoE2 in age [24]. ApoE4 may be adaptive, reducing mortality in highly infectious environments, with food scarcity and shorter lifespans [35]. However, as human environments became less septic, with food abundance and longer life expectancy, ApoE4 started to burden the arteries and brain, increasing the risk of diseases related to ageing [24]. The emergence of ApoE3 from ApoE4 putatively reflects the shift in human diet from a plant-based one to a meat-rich one, where genes adaptive to high meat consumption were, and still are vital to regulate increased cholesterol levels [36].

*Tissue expression.* The principal producers of ApoE are hepatocytes in the liver [37]. In the CNS, ApoE is primarily expressed in glia (non-neuronal cells of the brain that support the neurons) and particularly in astrocytes (metabolic homeostasis and neuronal communication modulators), followed by microglia (the brain immune cells) [38]. Each genotype is linked to different expression levels [13]. ApoE2 carriers seem to have higher cerebrospinal fluid (CSF) levels of ApoE, compared to ApoE4 carriers [39, 40].

### 1.3. Apolipoprotein E.

*Structure and function.* ApoE is a brain-specific lipid-binding glycoprotein of 299 amino-acids (34 kDa) that comprises several types of lipoproteins, i.e., chilomicra, intermediate density lipoprotein and very low density lipoprotein [13]. Its main function in the brain is the transport of lipids (mainly cholesterol) through membrane receptors [14]. Moreover, its isoforms have an effect on diverse cellular functions, e.g. synaptic integrity, glucose metabolism, A $\beta$  clearance, blood-brain barrier integrity and mitochondrial regulation [13]. How these relate to AD pathology will be elaborated in Section 1.3.

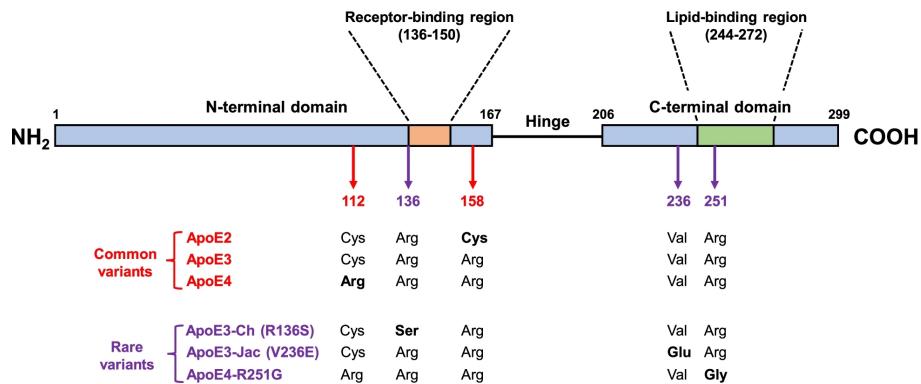
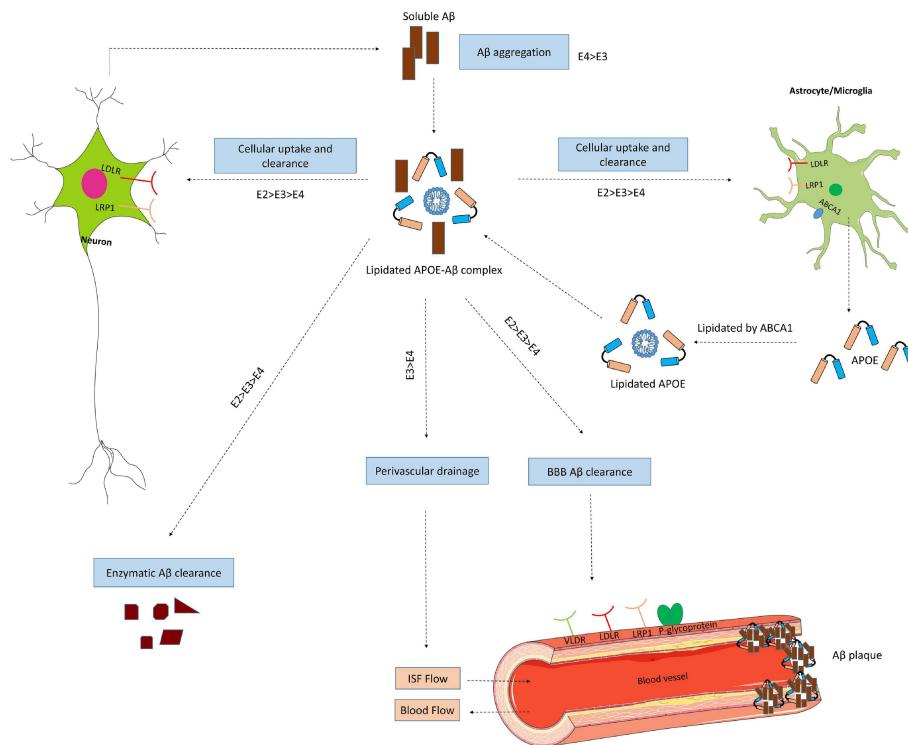


FIGURE 2. Linear representation of the ApoE protein. Three structural domains are highlighted: N-terminal, hinge and C-terminal domains. The different amino acids at positions 112 and 158 are shown per common alleles and amino acids at positions 136, 236 and 251 coded by rarer alleles. Source: “APOE targeting strategy in Alzheimer’s disease: lessons learned from protective variants”, Bu (2022) [41]

*ApoE isoforms.* The nuances in the amino acid composition of ApoE, specifically the presence of cysteine or arginine at positions 112 and 158, significantly impact its binding with lipids and receptors [24]. The most prevalent isoform, ApoE3, features cysteine at position 112 and arginine at position 158 [24], as shown in Fig. 2. ApoE2 has two cysteines, while ApoE4 has two arginines at these positions [24]. The C-terminal domain of ApoE (positions 273–299) is crucial for its lipidation specificity and efficiency [42].

*Lipidation nuances of ApoE isoforms.* For ApoE to exert its effects, it needs to be lipidated [13]. ApoE undergoes lipidation via ATP-binding cassette transporter A1 (ABCA1), a lipid efflux protein [43, 44]. The lipidation degree varies among ApoE isoforms, with ApoE4 exhibiting the least efficient lipidation [42, 45]. This discrepancy in lipidation has been linked to alterations in lipoprotein size and type, in that ApoE4 “prefers” large triglyceride-rich VLDL, while ApoE2 and ApoE3 have a higher affinity for phospholipid-rich HDL particles [46]. The lower affinity of ApoE4 for HDL particles leads to increased levels of unlipidated ApoE, resulting in its aggregation [47]. Moreover, ApoE4 fibrils are more neurotoxic than those of ApoE2 and ApoE3 [48].

Poor lipidation leads to poor ApoE recycling [24] (Fig. 3). The latter favors the entrapment of ABCA1 in endosomes, away from the cell membrane, thereby pooling cholesterol in the cell membrane rather than attaching it to HDL particles [49]. This increased cholesterol content in the cell membrane amplifies toll-like receptor 4 (TLR4) signaling in macrophages, activating NF $\kappa$ B and inducing an inflammatory gene response [24]. ApoE4 accumulation also sequesters the insulin receptor (IR) in endosomes, impacting cellular energy preferences [50]. This leads to a decrease in glucose utilization for ATP production (Fig. 4) and an increase in fatty acid oxidation [51].



**FIGURE 3.** Effects of ApoE isoforms on the metabolism and removal of A $\beta$ . A $\beta$  is primarily cleaved from amyloid precursor protein (APP). In the brain, ApoE is mainly expressed in astrocytes and microglia, and undergoes lipidation by ATP-binding cassette transporter A1 (ABCA1) to create lipoprotein particles. ApoE increases the accumulation and buildup of A $\beta$ , or promotes cellular uptake of A $\beta$  by astrocytes or microglia via endocytosis of the lipidated ApoE-A $\beta$  in an isoform-specific manner. This process involves several receptors, such as low-density lipoprotein receptor (LDLR) and low density lipoprotein receptor-related protein 1 (LRP1). ApoE also facilitates isoform-specific breakdown of A $\beta$  outside the cells. At the blood-brain barrier, soluble A $\beta$  is predominantly transported from the interstitial fluid (ISF) into the bloodstream via LRP1 and P-glycoproteins. Additionally, ApoE plays a role in the perivascular drainage of A $\beta$ . Insufficient clearance of A $\beta$  can lead to its accumulation in the brain tissue, contributing to the formation of A $\beta$  oligomers and amyloid plaques. Source: “APOE and Alzheimer’s Disease: From Lipid Transport to Physiopathology and Therapeutics”, Husain *et al.* (2021) [13]

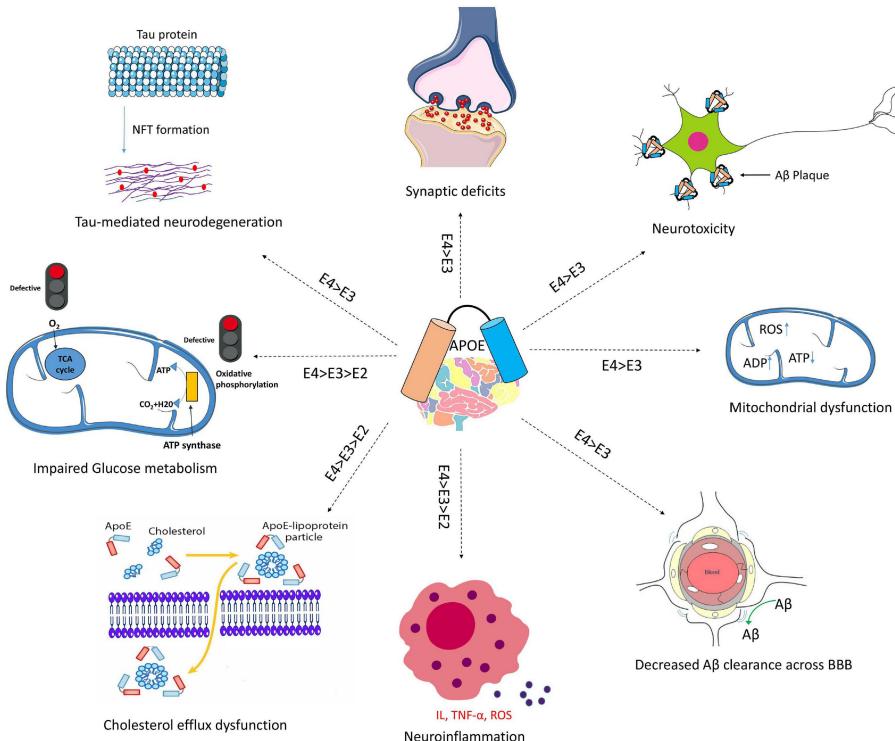
*Interplay between ApoE lipidation and Alzheimer’s disease pathophysiology.* As mentioned earlier, ApoE isoforms have differential pleiotropic effects on various cellular functions. ApoE4 induces a pro-inflammatory response, leading to the dysfunction of the blood-brain barrier, which in turn impairs cognitive functions [52–54]. Moreover, ApoE modulates the primary neuropathological hallmarks of AD: neuroinflammation, A $\beta$  plaques and tau tangles [13]. Evidence from human and transgenic mice studies reveals increased brain A $\beta$  and amyloid plaque loads in ApoE4 carriers, compared to ApoE3; with the lowest levels in ApoE2 carriers [55–57]. ApoE4 has a higher binding affinity for A $\beta$  which leads to impaired clearance, its intracellular aggregation and higher plaque loads [54] (Fig. 3). Additionally, high levels of ApoE4 in neurons remarkably increase tau protein phosphorylation, while high concentrations of ApoE3 do not seem to have

an effect [58–61]. Notably, ApoE directly inhibits phosphorylation of tau by glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ) [62]. An overview of the A $\beta$ -(in)dependent effects of ApoE is shown in Fig. 4.

**1.4. ApoE4-mediated metabolic changes in Alzheimer's disease.** Metabolism entails the repertoire of chemical reactions that keep living organisms alive. Metabolites –especially lipid [63–65]–, perceived as functional intermediates of AD development, are rigorously studied for biomarkers or targets for treatment [66].

*Measured in post-mortem brain tissue.* A metabolomic profiling of brain tissue, obtained *post-mortem* from AD patients and healthy controls showed pronounced impairments in sterol and sphingolipid levels in ApoE4 carriers with AD [67]. However, another *post-mortem* metabolomic analysis did not reveal nuances significantly correlated with ApoE4 [68], although they showed trends in increased cholesterol esters, unsaturated lipids, and sphingomyelin species.

*Measured in blood.* Transcriptomic and lipidomic analyses in humanized ApoE mice associated ApoE4 with decreased free fatty acid levels, many increased tricarboxylic acid (TCA) cycle metabolites, as well as changes in plasma levels of phosphatidylcholines and unsaturated fatty acids [69, 70]. A recent study on



**FIGURE 4.** Schematic overview of A $\beta$ -independent effects of ApoE in AD pathology. ApoE4 increases the phosphorylation of tau proteins, leading to the creation of tangles, inducing neurodegeneration, synaptic deficits and neurotoxicity. Moreover, ApoE4 is associated with decreased cholesterol efflux and neuroinflammation mediated by interleukin (IL), tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ), reactive oxygen species (ROS). In the mitochondria, ApoE4 impairs glucose metabolism and ATP production. Source: “APOE and Alzheimer’s Disease: From Lipid Transport to Physiopathology and Therapeutics”, Hu-sain *et al.* (2021) [13]

58 individuals found six downregulated plasma metabolites (including lysophospholipids and cardiolipin) in ApoE carriers [71]. Further, the plasma metabolome of the latter reveals a preference for aerobic glycolysis [72]. Significant correlations of ApoE genotype and sex with metabolites were observed, i.e. several phosphatidylcholines were found in a large study of more than 1500 individuals [23].

Perturbed serum metabolites associated with AD are aminoacids, amines [73, 74], cholesterly esters [65], sphingolipids [63, 66, 74–76], fatty acids [64, 73], glycerophospholipids [75, 77–79], phosphatidylcholines [80] and lipid peroxidation compounds [64]. These molecules are usually identified via high-throughput metabolomic pipelines (coupled with Mass Spectrometry (MS) detectors) that trace all compounds in a sample and result in high-dimensional data [81]. The latter often require advanced statistical methods e.g. projection to latent structures [79, 82] or graphical models [83] in order to extract putatively meaningful information. With such techniques, de Leeuw et al. discovered distinct serum metabolic signatures among AD patients-controls and those carrying at least one ApoE  $\epsilon_4$  allele [73], as they are shown in Fig. 5. The different metabolic profiles, however, among ApoE4 non-carriers remain obscure. The present data science approach shows ApoE4-mediated differentially expressed metabolites, potentially unveiling distinct pathways of metabolic deregulation in AD.

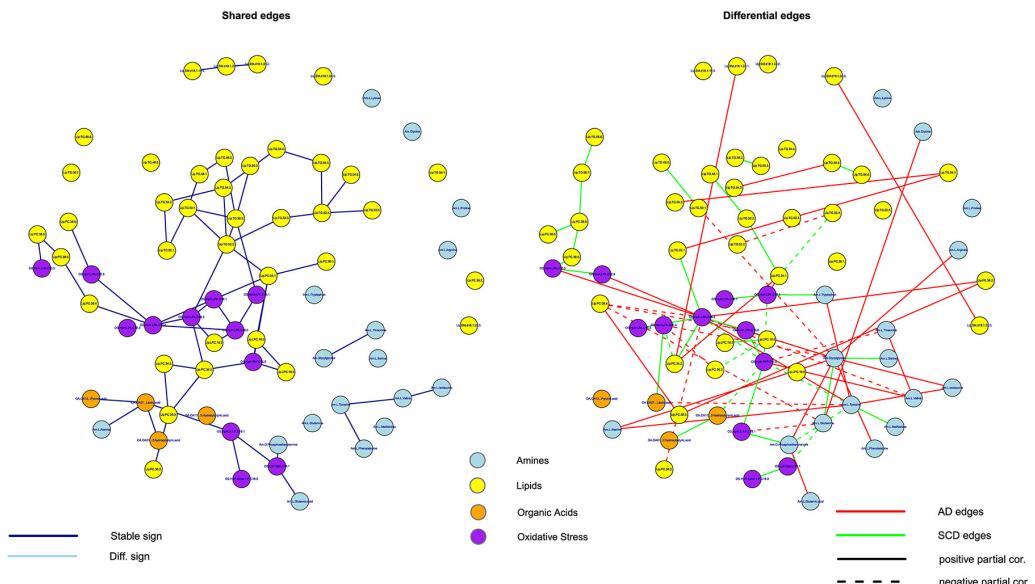


FIGURE 5. Mutual (left-hand panel) and distinct (right-hand panel) metabolic network topologies between ApoE4 carriers with AD and non-carriers with SCD, as published by de Leeuw et al.. Red edges represent links that are present exclusively in ApoE4 carriers. Green edges represent connections found in the SCD group without ApoE4. Solid edges represent positive partial correlations, while dashed edges represent negative partial correlations. Abbreviations: SCD, subjective cognitive decline. Source: “Blood-based metabolic signatures in Alzheimer’s disease”, De Leeuw *et al.* (2017) [73]

**1.5. Research Questions.** ApoE4 carriers –particularly females– experience metabolic disturbances and are at increased risk of LOAD. The mechanistic links, however, between ApoE4 dose, metabolism and AD development are not entirely known [64]. Tracking ApoE4 dose effects on serum metabolites can reveal metabolic perturbations preceding or co-existing with AD. Hence, in an effort to elucidate ApoE4-mediated changes in serum metabolites in AD and SCD, one could state the following research questions (RQ):

Are there mechanistic links between ApoE4 dose and serum metabolome in AD?

- (1) Are there ApoE4 dose effects on serum metabolite levels in AD?
- (2) How discriminatory are serum metabolites of ApoE4 status and/or AD?
- (3) How do the covariance network topologies of metabolites differ between ApoE4 carriers and non-carriers?

**1.6. Approach and Overview.** An introduction to the data used in this study is found in Section 2.1. A general overview of the software is found in Section 2.2, while the R session information is found in Appendix B. To facilitate statistical analysis, two new features were created for the first two research questions: "ApoE4Dose" (0, 1, 2) and "ApoE4AD" (4 possible phenotypes: AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4, SCD with at least 1 ApoE4), respectively, as described in Section 2.3.

The statistical methods applied to answer the research questions were adapted from De Leeuw *et al.*'s "Blood-based metabolic signatures in Alzheimer's disease" and are described in Section 2.4. To screen for ApoE4 dose effects on serum metabolites, two approaches were taken: a global test (correcting only for sex) and nested linear model comparison using ANCOVA F-tests (correcting for several factors: Table 1, except CSF markers).

To test the (added) classification potential of serum metabolites against ApoE4AD, several multi-class classification models were fitted. First, a benchmark multinomial logistic regression model was fitted using only clinical background data as predictors. Second, the full metabolomic panel was added on top of the clinical data in the same model and feature importance was calculated. Third, the metabolites were projected to a latent orthogonal space, where 6 meta-metabolites (accounting for around 30% of the variance) replaced the original metabolites. Finally, the meta-metabolites were fitted on top of the clinical data in a multinomial logistic regression model, a decision tree and an eXtreme gradient boosting (XGBoost) model. The discriminatory performance of the aforementioned models was then comprehensively evaluated and compared.

To visualise and compare the covariance network topologies of metabolites among ApoE4 carriers and non-carriers, the high precision matrices were first regularised with Ridge, sparsified and then pruned controlling the False Discovery Rate. Network topologies were plotted and their statistics were calculated and compared between the ApoE4 carriers and non-carriers using Wilcoxon Signed Rank test.

The results of the analysis are presented in Section 3 and discussed in Section 4. The study is concluded in Section 5

## 2. METHODS

**2.1. Subjects.** The data were collected from 120 AD patients and 127 SCD ( $n = 247$  in total) individuals recruited within Amsterdam Dementia Cohort [73, 84]. In this study two data sets were used: a targeted metabolomics panel and clinical background data, i.e. age at diagnosis, sex, smoking status, alcohol consumption, hypertension (and medication), hyperlipidemia (and medication), anticoagulant medication, antidepressants, mean arterial pressure (MAP) and body mass index (BMI) (see Table 1). The metabolomics set contains  $p = 230$  metabolites (amines, organic acids, lipids and oxidative stress compounds). The methodology for the metabolomic analysis and ApoE genotyping can be found at de Leeuw et al.’s Blood-based metabolic signatures in Alzheimer’s disease [73]: SMT1. The data were cleaned as described in the same article. The resulting datasets for AD and SCD are high-dimensional, in the sense that they contain more variables than observations ( $p > n$ ). Another particularity of the data is the covariance and collinearity of the variables. Therefore, appropriate measures need to be taken to prevent model over-fitting –the algorithm being unable to distinguish signal from noise and fitting the latter– and to correct for spurious correlations.

TABLE 1. Clinical background characteristics used as control variables of ApoE4 (dose) effects.

	Clinical feature	Type
Anthropometric	Age	Discrete
	Sex	Binary
Intoxications	Smoking (past, current, no)	Nominal
	Alcohol	Binary
Comorbidities	Hypertension	Binary
	Diabetes mellitus	Binary
	Hypercholesterolemia	Binary
Medication	Cholesterol lowering	Binary
	Antidepressants	Binary
	Antiplatelet	Binary
CSF markers*	$\text{A}\beta_{42}$	Continuous
	tau	Continuous
	p-tau	Continuous

\* were only used in the ApoE4AD status multi-class classification

**2.2. Data management.** The FAIR principles for data management and stewardship in science were published by Wilkinson *et al.* in 2016 [85]. FAIR stands for Findable, Accessible, Interactive, and Reusable data; the intention is to create and use data that are well-documented and reproducible. These principles were considered at every step of the study and implemented when applicable. The statistical analysis was performed in R (version 4.3.2), and the current report was written in L<sup>A</sup>T<sub>E</sub>X(Tex Live version 2023). All files are stored in a private Github repository –with git as version control system (VCS).

**2.3. Feature Engineering.** The information of the ApoE genotypes is valuable, and it might be interesting to screen for metabolic nuances between them. However, the genotypes were not equally represented in the data and hence, testing for differences among them would not be possible. The following two subsections describe the features ApoE4Dose and ApoE4AD that were created to balance the genotypes.

TABLE 2. Observed ApoE genotype (top part), allele (middle part) and ApoE4 allelic (bottom part) counts in AD and SCD

		AD (%)	SCD (%)	Total (%)
<i>genotypes</i>	$\varepsilon_2/\varepsilon_2$	2 (1.7)	0 (0.0)	2 (0.8)
	$\varepsilon_2/\varepsilon_3$	15 (12.5)	3 (2.4)	18 (7.3)
	$\varepsilon_2/\varepsilon_4$	5 (4.2)	2 (1.6)	7 (2.8)
	$\varepsilon_3/\varepsilon_3$	69 (57.5)	37 (29.1)	106 (42.9)
	$\varepsilon_3/\varepsilon_4$	26 (21.7)	59 (46.5)	85 (34.4)
	$\varepsilon_4/\varepsilon_4$	3 (2.5)	26 (20.5)	29 (11.7)
<i>alleles</i>	$\varepsilon_2$	24 (10.0)	5 (2.0)	29 (6.0)
	$\varepsilon_3$	179 (74.6)	136 (53.5)	315 (64.0)
	$\varepsilon_4$	37 (15.4)	113 (44.5)	150 (30.0)
$\varepsilon_4$	1x*	31 (25.8)	61 (48.0)	92 (37.2)
	2x*	3 (2.5)	26 (20.5)	29 (11.7)
	$\geq 1x^\dagger$	34 (28.3)	87 (68.5)	121 (49.0)
	No*†	86 (71.7)	40 (31.5)	126 (51.0)
	Total	120 (49.0)	127 (51.0)	247

\* rows used for ApoE4Dose

† rows used for ApoE4AD

*ApoE4 dose effects.* De Leeuw *et al.* divided the subjects into two classes: those carrying at least one  $\varepsilon_4$  allele, and  $\varepsilon_4$  non-carriers. However, in order to study the  $\varepsilon_4$  dose effects, the genotypes can be categorized into groups, based on the number of  $\varepsilon_4$  alleles they carry: 0, 1 or 2. The number of ApoE4 allele doses is shown in the bottom part of Table 2.

*Classification of ApoE4 and AD status.* In order to incorporate ApoE4 status, as well as the diagnosis of AD, a four-level feature (AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4, SCD with at least 1 ApoE4, "ApoE4AD") was created, as shown in 2.

#### 2.4. Statistical Analysis.

2.4.1. *ApoE4 dose effects on serum metabolite levels in AD.* To test if the number of ApoE4 alleles have an effect on mean metabolite levels in AD, two methods were applied: a global test and nested linear model comparison with ANCOVA.

**Global Test** The concept of a global test was first introduced by Simon *et al.*, and proposed an approach based on permutations to cater to the high dimensionality of gene expression data. The R package *globaltest* [87–89] developed by Goeman *et al.* features a multinomial logistic regression model, fitting genes to predict clinical or biological group membership (number of ApoE4 alleles in this case). This method is also appropriate for other types of -omics data, such as metabolomics in this study [89]. The null hypothesis is that metabolite levels are independent of the ApoE4 dose X, i.e.  $H_0 : P(Y|X) = P(Y)$ , where  $X \in \mathbb{R}^{n \times p}$ . The test statistic under  $H_0$  follows, asymptotically, a normal distribution. The *gt* function of the package was used to screen for nuances in metabolite levels on the number of ApoE4 alleles, correcting for sex. To assess ApoE4 dose effects correcting for clinical data nested linear model comparison was performed, as described in the next section.

**Nested linear models** An linear model with the Least Squares estimation creates a regression using categorical and numerical variables, while minimizing the Residual Sum of Squares (RSS) [90],

$$\text{RSS} = \sum (y - \hat{y})^2$$

where  $y$  is the response and  $\hat{y}$  is its regression estimate. Two models were fitted for each metabolite, a full and a nested model. The dependent variable in each model was a metabolite; the nested model (1), has only clinical variables (Table 1 except CSF markers) while the full model (2), features the clinical variables, plus the number of ApoE4 alleles (0, 1 or 2 -nominal) as explanatory variables. Let  $y_j$  represent the  $j$ -th metabolite,  $x_k$  the  $k$ -th clinical variable, and  $D_{\varepsilon_4}$  the ApoE4 dose; the nested models then are:

$$(1) \quad y_j = \beta_0 + \sum_{k=1}^m \beta_k x_k + \epsilon$$

$$(2) \quad y_j = \beta_0 + \sum_{k=1}^m \beta_k x_k + \beta_{m+1} D_{\varepsilon_4} + \epsilon$$

For every metabolite  $j = 1, \dots, p$  the hypothesis test is

$$H_0 : \beta_{m+1} = 0 \text{ vs } H_\alpha : \beta_{m+1} \neq 0$$

The test statistic is ANCOVA F-test:

$$F = \frac{\text{SSReg}_{full} - \text{SSReg}_{nested}/df_{full} - df_{nested}}{\text{RSS}_{nested}/n - (m + 2)}$$

Under  $H_0$   $F$  follows an  $F_{1,n-(m+2)}$  distribution [90]. The null hypothesis is rejected for large values of  $F$ .

This implies  $p$  hypothesis tests, which qualifies as multiple testing (the probability of incorrectly rejecting  $H_0$ , type I error, increases monotonically with every additional test). A method to treat the latter is controlling False Discovery Rate [91], that is controlling the expected ratio of incorrectly rejected  $H_0$  hypotheses, globally e.g. at an  $\alpha$  of 0.05. With the Benjamini-Hochberg's approach, first, the p-values are sorted in ascending order. Then, for every p-value  $j$ ,  $\alpha$  is multiplied by  $j$  over the total number of tests [91]  $m = 230$  in the AD group in this study. In other words, after adjustment, a null hypothesis  $j$  may be rejected only if its associated F-test's p-value is less or equal to a fraction  $(j/m)$  of  $\alpha$ .

---

**Algorithm 1:** Benjamini–Hochberg's procedure to control FDR

---

- 1 Specify  $\alpha$ , the level at which to control the FDR.
- 2 Compute p-values,  $p_1, \dots, p_m$ , for the  $m$  null hypotheses  $H_{01}, \dots, H_{0m}$ .
- 3 Order the  $m$  p-values so that  $p(1) \leq p(2) \leq \dots \leq p(m)$ .
- 4 Define

$$L = \max\{j : p(j) \leq \frac{j}{m}\alpha\}$$

- 5 Reject all null hypotheses  $H_{0j}$  for which  $p_j \leq p(L)$ .
- 

To test for ApoE4 dose effects on each metabolite in AD and SCD,  $2m = 460$  model comparisons were needed in total, hence a function was created to iterate over the metabolites, fit the nested models, compare them using ANCOVA F-tests, store and adjust the p-values, filter those below .05, then consolidate the coefficients of the meaningful full models and the p-values of their t-tests in a table and display it. To decrease run time, as well as harness the power of multiple CPU cores, furr's `future_map` function was used for parallel iterations.

**2.4.2. Classification of ApoE4 and AD status.** In machine learning, a class denotes a group of objects that share common characteristics, such as having AD or ApoE4 [92]. Classification, in this context, denotes training a (supervised) learning algorithm on labelled data (containing their class) [92]. The classifier learns patterns in the data and is then able to predict class membership for unknown data [92].

**Bias-Variance trade-off** The degree to which a user can understand and interpret the prediction or decisions made by a statistical model is defined as *interpretability* [93]. It is of interest in this study to find the optimal balance between the performance of a model and its interpretability. The *bias-variance trade-off* was formally introduced by Geman *et al.* and refers to the trade-off between the accuracy (opposite of bias) and precision (opposite of variance) of a prediction. It also refers to the trade-off between model flexibility (or complexity) and interpretability [94]. One may consider this trade-off during model and evaluation method selection, as some impose more bias or variance than others.

**Multi-class classification models** The R package `caret` streamlines the training and comparison of classification and regression models, offering broad model training options and feature importance estimation [95]. The functions `trainControl` and `train` were used to fit the models. A sampling method used to deal with the unbalanced classes was SMOTE (Synthetic Minority Over-Sampling Technique), as implemented by the package `DMwR2` [96].

Considering interpretability, Multinomial Logistic Regression (MLR) is inherently interpretable. Let  $y$  a response with  $K$  classes,  $k \in N, [1, 4]$  representing the  $k$ -th class of ApoE4AD and  $\beta_{kj}$  its set of coefficients,  $\beta_{lj}$  the coefficients of the rest of classes for  $j$ -th metabolite, then a multinomial logistic regression model calculated the probability

$$P(y = k | X = x) = \frac{e^{\sum_{j=1}^p \beta_{kj} x_j}}{\sum_{l=1}^K \sum_{j=1}^p e^{\beta_{lj} x_j}}$$

When  $p > n$ , the coefficient estimation method has low bias and high variance, in that small changes in the training data can result in very different coefficient estimates [97].

Regularization trades off a small increase in bias for a great decrease in variance, by shrinking the low coefficients towards zero [97]. Ridge or L2 regularization [98] shrinks the coefficients without setting them to exactly zero [98]

$$\text{RSS} + \lambda \sum_{j=1}^p \beta_j^2$$

where  $\lambda \geq 0$  is a tuning parameter that balances the coefficient shrinking effect. The function with the same name of the package `nnet` fits a shallow neural network (with a single hidden layer, but allowing skip-layer connections) [99]. For the  $i$ -th observation, it calculates the weights using a Least Squares estimation of the negative conditional log-likelihood that it belongs to the  $k$ -th class

$$E = \sum_i \sum_k -y_{ik} \log \hat{y}_{ik}, \quad \hat{y}_{ik} = \frac{e^{\sum_{j=1}^p \beta_{kj} x_{ij}}}{\sum_{l=1}^K \sum_{j=1}^p e^{\beta_{lj} x_{ij}}}$$

where  $y_{ik}$ , the true class will be exactly one and the others all zero [99]. It regularises the fit via weight decay, a Ridge-like penalty that uses the sum of squares of the weights [99].

Another method to treat multi-collinearity and high dimensionality is a 2-stage maximum likelihood (ML) Factor Projection to a Latent Orthogonal space, such as the one the package `FMradio` [82] performs. In the 1st stage, a L2-regularised ML estimation is used to filter out redundant features from the data matrix. In the second stage, the aforementioned matrix is projected to an orthogonal space where the features are replaced by -fewer- factors that explain their covariance. One can then use the produced factor scores as predictors in any model.

Decision Trees (DT) are inherently interpretable, non-parametric models, that fit well large and complicated data sets. They have a tree-like structure that splits the data into branches and leaves (nodes) [100]. The `rpart2` [101] function of `rpart` was used.

Boosting models, are ensemble models that fit several weak learners (such as linear/logistic regression or DTs) sequentially, reweighing the data, and take their weighted majority vote [102, 103]. Despite Boosting tends to outperform DTs, it often operates as a *black box* and is poorly interpretable. The state-of-the-art eXtreme gradient boosting (`xgbTree`) of the package `xgboost` [104] was used.

---

**Algorithm 2:** Multi-class classification of ApoE4 and AD status pipeline

---

- 1 Fit clinical data only in multinomial logistic regression
  - 2 Fit clinical data + 230 serum metabolites in multinomial logistic regression
  - 3 Fit clinical data + 6 meta-metabolites in multinomial logistic regression
  - 4 Fit clinical data + 6 meta-metabolites in a Decision Tree
  - 5 Fit clinical data + 6 meta-metabolites in an XGBoost
  - 6 Evaluate performance and compare
- 

**Implementation & Evaluation** First, a benchmark model was created fitting the clinical background data to predict ApoE4AD in a penalised multinomial logistic regression model. Second, the 230-metabolite panel was fitted on top of the clinical background data in a penalised multinomial logistic regression model ( $\lambda = 9.187724$ , obtained from repeated 10-fold Cross-Validation (CV)) and feature importance was calculated. Then the full metabolite panel was projected into 6 latent factors (meta-metabolites), explaining around 30% of their variance. The 6-factor metabolite projection was then fitted on top of the clinical data in a penalised multinomial logistic regression ( $\lambda = 0$ ), a decision tree and an XGBoost model. The aforementioned models were hyperparameter tuned over a grid of values and the best was selected using repeated (100 times) 10-fold CV.

Model performance was comprehensively assessed and compared, with the repeated 10-fold CV-obtained receiver-operating characteristics (ROC) curves and their respective Area Under the Curve AUC using the `pROC` package [105] and other metrics such as Accuracy, Precision, Recall and F1-score from `caret`'s `confusionMatrix`. Feature importance was estimated using `caret`'s `varImp`.

**2.4.3. Metabolite Covariance Network Analysis.** Network science offers a unifying framework for data and system representation, applicable to any domain [106]. A network, in an abstract sense, consists of nodes connected with links, also referred to as edges. In data science, a network whose nodes represent random features, whose joint probability distribution is defined by the ensemble of their edges is called *graphical model* [83]. A metabolomic covariance correlation network represents the ensemble of metabolites based on their covariance, showing nuances among the samples that non-graphical statistical methods on individual metabolites may fail to detect [107]. It may provide insights into correlated metabolites that do not belong in the same metabolic pathway [107].

A *Gaussian graphical model* (GGM) is an undirected graph that represents the conditional independence properties of the features [108]. The statistic employed by GGMs is the partial correlation which also adjusts for indirect correlation, i.e. two metabolites are correlated with a third one and are shown correlated with each other [109]. For instance, let  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  be a GGM consisting of a set  $\mathcal{V}$  of  $p$  vertices, corresponding to random features  $Y_1, \dots, Y_p$  with joint probability distribution  $P \sim N_p(\mathbf{0}, \Sigma)$ , and set of edges  $\mathcal{E}$ , such that for all pairs  $\{Y_i, Y_j\}$  with  $i \neq j$ :

$$\Sigma_{ij}^{-1} = (\Omega_{ij}) = 0 \iff Y_i \perp\!\!\!\perp Y_j \mid \{Y_k : k \neq i, j\} \iff (i, j) \notin \mathcal{E}$$

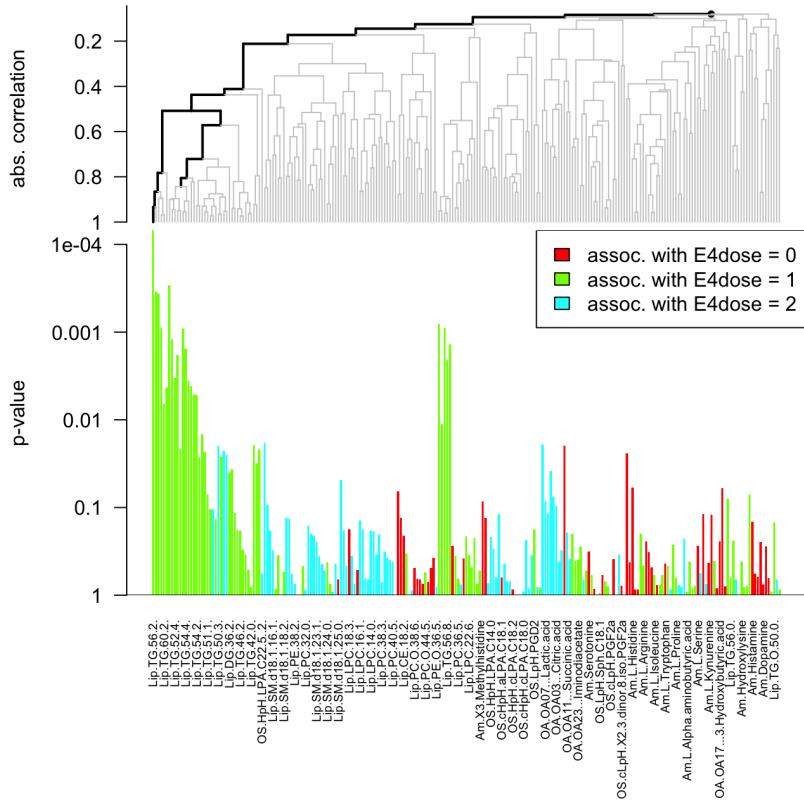
In natural language, a zero value in the inverse covariance matrix (usually referred to as precision matrix  $\Omega$ ) implies that the respective random features are independent, given the rest of features, and they are not connected by an undirected edge  $((i, j) \notin \mathcal{E})$  [83]. In this study, the package `rags2ridges` [83] was used to generate the feature covariance matrices of ApoE4 carriers and non-carriers with AD, compute their precision matrices, regularise them and represent them in GGMs. Communities of coregulated metabolites were identified using the community search algorithm by Girvan-Newman [110] as implemented by `rags2ridges`.

### 3. RESULTS

The ApoE genotype frequencies in AD and SCD are shown in Table 2. Cumulatively, the most abundant allele is  $\varepsilon_3$  (64%), followed by  $\varepsilon_4$  (30%) and  $\varepsilon_2$  (6%). In AD, the allelic frequencies were 74.6% for  $\varepsilon_3$ , followed by 15.4% for  $\varepsilon_4$  and 10% for  $\varepsilon_2$ . In the SCD group,  $\varepsilon_3$ ,  $\varepsilon_4$  and  $\varepsilon_2$  showed relative frequencies of 53.5%, 44.5% and 2%. The most common genotype was  $\varepsilon_3/\varepsilon_3$ , followed by  $\varepsilon_3/\varepsilon_4$  and  $\varepsilon_4/\varepsilon_4$ .

#### 3.1. ApoE4 dose effects on serum metabolite levels in AD.

3.1.1. *Global Test.* Testing for ApoE4 dose-effect on serum metabolites of AD patients, correcting for sex ( $H_0$ : ApoE4 dose has no effect on mean metabolite levels,  $H_a$ : it has an effect), showed a significant global difference in metabolites ( $p = 0.017$ ). The most significantly affected metabolites are triglycerides and diglycerides (FDR-adjusted  $p$ -value  $<0.05$ ) See Table 3 and Fig. 6. In SCD individuals, the global test showed no significant differences among ApoE4 doses ( $p=0.544$ ). Generally, fewer metabolites were altered in this group and none of the associations remained significant after adjusting for FDR. Interestingly, the affected metabolites are different between AD and SCD.



**FIGURE 6.** Covariates plot showing the metabolites affected by the number of ApoE4 alleles in AD

TABLE 3. Metabolites affected by ApoE4 dose, correcting for sex as per globaltest [89]

	Class	Metabolite	Inheritance	Assoc. with	p-value	FDR
AD	Lipid	TG (56:2)	0.046	1 ApoE4	<0.001	0.016
	Lipid	TG (58:1)	0.117	1 ApoE4	<0.001	0.021
	Lipid	DG (36:3)	0.19	1 ApoE4	<0.001	0.021
	Lipid	TG (56:3)	0.244	1 ApoE4	<0.001	0.021
	Lipid	TG (52:3)	0.424	1 ApoE4	0.001	0.031
	Lipid	TG (54:5)	0.48	1 ApoE4	0.001	0.027
	Lipid	TG (58:2)	0.722	1 ApoE4	0.001	0.027
	Lipid	TG (54:4)	0.761	1 ApoE4	0.002	0.033
	Lipid	TG (58:9)	1	1 ApoE4	0.001	0.027
	Lipid	TG (56:7)	1	1 ApoE4	0.001	0.027
	Lipid	TG (58:8)	1	1 ApoE4	0.001	0.032
	Lipid	TG (54:6)	1	1 ApoE4	0.002	0.035
	Lipid	TG (56:8)	1	1 ApoE4	0.002	0.037
	Lipid	TG (52:4)	1	1 ApoE4	0.003	0.055
	Lipid	TG (54:3)	1	1 ApoE4	0.004	0.055
	Lipid	TG (56:6)	1	1 ApoE4	0.004	0.059
	Lipid	TG (56:1)	1	1 ApoE4	0.004	0.059
	Lipid	TG (52:2)	1	1 ApoE4	0.005	0.063
	Lipid	TG (54:2)	1	1 ApoE4	0.005	0.063
SCD	Lipid	TG (60:2)	1	1 ApoE4	0.007	0.077
	Lipid	TG (58:10)	1	1 ApoE4	0.011	0.124
	Lipid	TG (51:3)	1	1 ApoE4	0.015	0.154
	Lipid	SM (d18:1/18:1)	1	2 ApoE4	0.018	0.169
	Organic acid	2-ketoglutaric.acid	1	2 ApoE4	0.019	0.169
	Lipid	TG (52:0)	1	1 ApoE4	0.019	0.169
	Lipid	TG (50:3)	1	2 ApoE4	0.02	0.169
	Organic acid	Uracil	1	no ApoE4	0.02	0.169
	Lipid	TG (52:5)	1	1 ApoE4	0.021	0.174
	Lipid	TG (54:0)	1	1 ApoE4	0.022	0.174
Ox. Stress	Lipid	TG (50:2)	1	2 ApoE4	0.022	0.174
	Lipid	TG (51:2)	1	1 ApoE4	0.023	0.174
	Aminoacid	L-Glutamine	1	no ApoE4	0.024	0.174
	Lipid	TG (50:1)	1	2 ApoE4	0.025	0.174
	Lipid	TG (50:4)	1	1 ApoE4	0.026	0.176
	Lipid	TG (52:1)	1	1 ApoE4	0.027	0.176
	Lipid	TG (54:1)	1	1 ApoE4	0.031	0.203
	Lipid	TG (50:0)	1	1 ApoE4	0.037	0.229
	Organic acid	Malic.acid	1	2 ApoE4	0.038	0.234
	Lipid	DG (36:2)	1	1 ApoE4	0.039	0.235
CE	Ox. Stress	PAF (16:0)	1	2 ApoE4	0.049	0.283
	Lipid	LPC (20:5)	1	no ApoE4	0.012	0.828
	Lipid	SM (d18:1/23:0)	1	1 ApoE4	0.015	0.828
	Aminoacid	L-Tryptophan	1	no ApoE4	0.024	0.828
	Aminoacid	Putrescine	1	1 ApoE4	0.028	0.828
	Aminoacid	Glycine	1	1 ApoE4	0.035	0.828
	Lipid	CE (18:2)	1	1 ApoE4	0.038	0.828
	Aminoacid	SM (d18:1/22:0)	1	1 ApoE4	0.039	0.828
	Ox. Stress	LPA (20:5)	1	no ApoE4	0.040	0.828
	Lipid	PC (36:3)	1	1 ApoE4	0.043	0.828
DG	Organic acid	Succinic acid	1	1 ApoE4	0.043	0.828
	Amine	Citrulline	1	1 ApoE4	0.043	0.828

CE: Cholestryler ester, DG: Diglyceride, LPA: Lyso-sphingolipid LPC: Lysophosphatidylcholine, PAF: Platelet activating factor, SM: Sphingomyelin, TG: Triglyceride

**3.1.2. Nested Linear Models.** Several metabolites from all classes were altered by ApoE4 dose, both in AD and SCD. However, after controlling FDR, none of the effects remain significant at  $\alpha = 0.05$ . Among AD patients, metabolites showing trends of a positive effect were several triglycerides, diglycerides, putrescine, 2-ketoglutaric acid, lysophosphatidylcholin, platelet activating factor (16:0) and lyso-phosphatidic acid (18:0) (Table 4). Among individuals with SCD, lipid metabolites were not affected as much as in the AD group, with only two sphingomyelin species showing a difference. Aminoacids L-serine, tryptophan, glycine, tryptophan, L-homoserine, putrescine showed trends of effect in this group. L-Tryptophan is negatively associated with ApoE4 dose (at 1x and 2x ApoE4), while L-serine, glycine and L-homoserine are negatively associated only with ApoE4 homozygotes. (Table 4).

TABLE 4. ApoE4 dose effects on serum metabolites in AD and SCD: results from nested linear model comparison. Full model: clinical background variables and number of ApoE4 alleles, Nested model: clinical background variables only. Colour range green-red reflects trends in negative or positive correlations between ApoE4 dose and metabolites.

	Metabolite	P(>F)	FDR	No $\varepsilon_4$		1x $\varepsilon_4$		2x $\varepsilon_4$	
				Coef.	P(>t)*	Coef.	P(>t)*	Coef.	P(>t)*
AD	Lip.TG (52:3)	0.001	0.156	-1.4	0.818	2.6	0.004	3.7	0.001
	Lip.TG (52:4)	0.003	0.156	0.6	0.888	1.6	0.008	2.4	0.002
	DG (36:3)	0.002	0.156	0.0	0.631	0.0	0.012	0.0	0.001
	OS.HpH.PAF (16:0)	0.002	0.156	34.8	<0.001	2.6	0.017	4.6	0.001
	Lip.TG (52:2)	0.006	0.255	-5.1	0.469	2.1	0.03	3.7	0.002
	Lip.TG (54:5)	0.01	0.373	1.9	0.496	0.9	0.016	1.3	0.006
	Lip.TG (50:2)	0.012	0.398	-4.5	0.297	0.9	0.141	2.2	0.003
	Lip.TG (50:1)	0.018	0.45	-2.3	0.539	0.7	0.179	1.8	0.005
	Lip.TG (54:4)	0.02	0.45	3.5	0.318	1.2	0.015	1.4	0.02
	Lip.TG (54:6)	0.017	0.45	-0.4	0.795	0.5	0.027	0.7	0.009
	Am.Putrescine	0.029	0.515	0.0	<0.001	0.0	0.035	0.0	0.017
	OA.2.ketoglutaric.acid	0.026	0.515	0.0	0.366	0.0	0.953	0.0	0.014
	Lip.TG (50:3)	0.028	0.515	-1.2	0.668	0.6	0.127	1.3	0.008
	Lip.TG (52:5)	0.042	0.538	0.5	0.789	0.4	0.134	0.7	0.014
	Lip.TG (56:7)	0.042	0.538	-2.4	0.095	0.5	0.015	0.4	0.105
SCD	Lip.TG (56:8)	0.044	0.538	-1.4	0.055	0.2	0.014	0.2	0.151
	Lip.TG (58:9)	0.042	0.538	-0.4	0.068	0.1	0.013	0.0	0.41
	LPC (16:0)	0.033	0.538	4.2	0.028	0.3	0.237	0.9	0.009
	OS.HpH.LPA (18:0)	0.044	0.538	0.5	0.038	0.0	0.178	0.1	0.013
	Am.L-Serine	0.002	0.285	4.7	<0.001	0.2	0.115	-1.1	0.003
	Am.L-Tryptophan	0.002	0.285	3.7	0.004	-0.3	0.047	-1.4	0.002
	Am.Glycine	0.006	0.45	4.3	0.001	0.4	0.015	-0.9	0.052
SCD	Am.L-homoserine	0.047	0.931	0.0	0.001	0.0	0.723	0.0	0.014
	Am.Putrescine	0.045	0.931	0.0	0.969	0.0	0.013	0.0	0.927
	Lip.SM (d18:1/22:0)	0.043	0.931	3.3	0.002	0.3	0.015	0.3	0.448
	Lip.SM (d18:1/23:0)	0.029	0.931	1.2	0.01	0.1	0.012	0.2	0.277

\* Not corrected for multiple testing.

Am: Aminoacid, Lip: Lipid, DG: Diglyceride, LPA: Lyso-sphingolipid LPC: Lysophosphatidyl-choline, SM: Sphingomyelin, TG: Triglyceride, OA: Organic Acid, OS: Oxidative Stress compound, PAF: Platelet activating factor

**3.2. Classification of ApoE4 and AD status.** All classification models had a multi-class ROC AUC above 0.8 and accuracies significantly higher than the No-Information Rate ( $p < 2.2e^{-16}$ ) (Table 1) in predicting AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4, SCD with at least 1 ApoE4. The worst-performing model was Multinomial Logistic Regression fitting the clinical data only, while the best-performing one was XGBoost fitting the clinical data and the 6 meta-metabolites. Adding serum metabolite information (either the full 230-metabolite matrix or its 6-factor projection) seems to slightly increase the discriminatory power of the models. Notably, looking at the individual ROC curves (Fig. 7, 8 and Appendix A), as well as the confusion matrices (Fig. 9 and 10) all models were able to discriminate better among certain classes (AD+E4 vs. SCD+E4, AD+E4 vs. SCD, AD-E4 vs. SCD+E4 and AD-E4 vs. SCD-E4) compared to others (AD+E4 vs. AD-E4 and SCD+E4 vs. SCD-E4).

TABLE 5. Performance metrics of multi-class classification of ApoE4 and AD status (AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4, SCD with at least 1 ApoE4), obtained from 10-fold CV repeated 100 times.

Model	Fitting	AUC	Accuracy (95/% CI)	P(NIR>Acc.)	NIR
MLR	Clinical features only	0.814	0.4807 (0.4744, 0.4869)	$< 2.2e^{-16}$	0.3522
	Clinical features + 230 metabolites	0.834	0.5753 (0.5692, 0.5815)	$< 2.2e^{-16}$	
		0.818	0.5313 (0.525, 0.5375)	$< 2.2e^{-16}$	
Decision Tree	Clinical features + 6 latent factors	0.818	0.5082 (0.502, 0.5145)	$< 2.2e^{-16}$	0.3522
	XGBoost	0.836	0.5944 (0.5882, 0.6005)	$< 2.2e^{-16}$	

AUC: Area Under the (ROC) Curve, MLR: Multinomial Logistic Regression, NIR: No-Information Rate

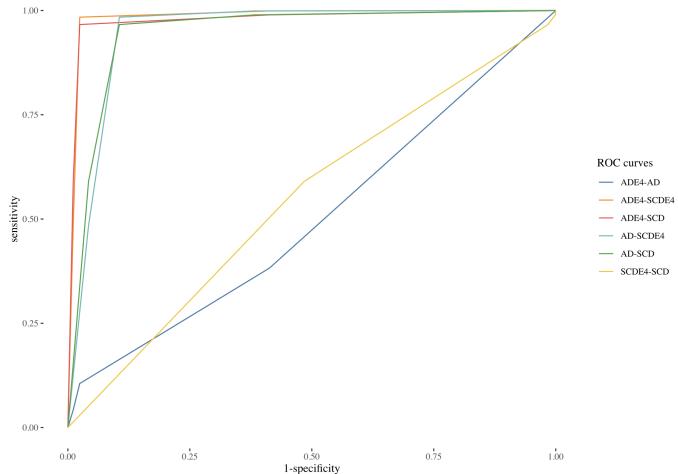


FIGURE 7. ROC curves of Multinomial Logistic Regression fitting the 230 metabolites on top of clinical background variables, obtained from repeated (100 times) 10-fold CV. AD: AD without ApoE4, ADE4: AD with at least 1 ApoE4, SCD: SCD without ApoE4, SCDE4: SCD with at least 1 ApoE4

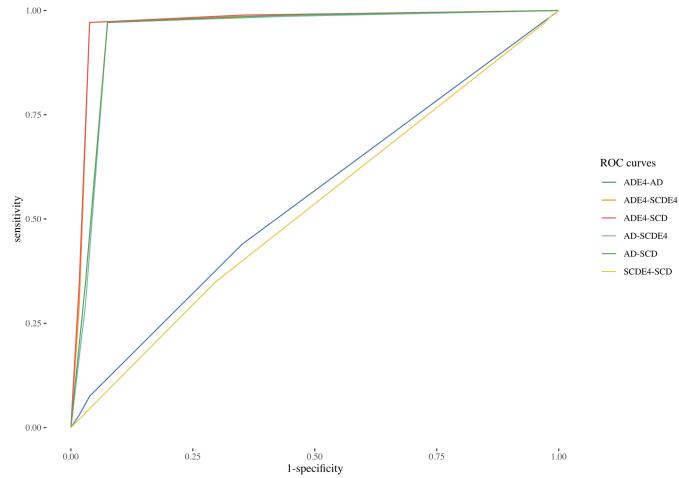


FIGURE 8. ROC curves of Multinomial Logistic Regression fitting the 6 ML-estimated latent factors on top of the clinical background variables, obtained from repeated (100 times) 10-fold CV. AD: AD without ApoE4, ADE4: AD with at least 1 ApoE4, SCD: SCD without ApoE4, SCDE4: SCD with at least 1 ApoE4

		Target			
		SCDE4	SCD	ADE4	AD
Prediction	SCDE4	23.6% 5823 66.9%	9.5% 2337 58.4%	0.2% 60 1.8%	0% 4 0%
	SCD	11.4% 2827 32.5%	6.7% 1662 41.5%	0% 6 0.2%	0% 4 0%
ADE4	SCDE4	0.2% 44 0.5%	0% 1 0%	3.9% 963 28.3%	11.5% 2829 32.9%
	AD	0% 6 0.1%		9.6% 2371 69.7%	23.3% 5763 67%

FIGURE 9. Confusion matrix showing the true (Target) and predicted (Prediction) classes of Multinomial Logistic Regression fitting the 230 metabolites on top of clinical background variables, obtained from repeated (100 times) 10-fold CV. AD: AD without ApoE4, ADE4: AD with at least 1 ApoE4, SCD: SCD without ApoE4, SCDE4: SCD with at least 1 ApoE4.

		Target			
		SCDE4	SCD	ADE4	AD
Prediction	SCDE4	22% 5427 62.4%	9.7% 2387 59.7%	0.5% 131 3.9%	0.3% 65 0.8%
	SCD	12.5% 3086 35.5%	6.3% 1559 39%	0% 6 0.2%	0.1% 24 0.3%
ADE4	SCD	0.7% 184 2.1%	0.1% 35 0.9%	5.1% 1263 37.1%	8.4% 2079 24.2%
	AD	0% 3 0%	0.1% 19 0.2%	8.1% 2000 58.8%	26% 6432 74.8%

FIGURE 10. Confusion matrix showing the true (Target) and predicted (Prediction) classes of XGBoost fitting the 6 ML-estimated latent factors on top of the clinical background variables, obtained from repeated (100 times) 10-fold CV. AD: AD without ApoE4, ADE4: AD with at least 1 ApoE4, SCD: SCD without ApoE4, SCDE4: SCD with at least 1 ApoE4

TABLE 6. Features with the highest feature importance scores in demarcating AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4 or SCD with at least 1 ApoE4 with a Multinomial Logistic Regression model fitting clinical background data and 230 metabolites.

Class	Feature	Overall (%)
Amine	Putrescine	100.00
Ox. stress	HpH.Spha.1.P.C18.0	97.43
Organic acid	Uracil	91.78
Lipid	TG (56:0)	88.86
Organic acid	3.Hydroxybutyric.acid	87.29
Clinical	Cholesterol medication	84.54
Amine	Sarcosine	83.54
Lipid	PE.O. (38:5)	81.60
Ox. stress	HpH.LPA.C20.5	72.43
Amine	L-Tryptophan	71.34
Lipid	SM (d18:1/20:1)	70.52
Amino acid	Glutathione	69.96
Amino acid	L-Serine	69.49
Amino acid	L-Histidine	69.41
Ox. stress	HpH.LPA.C18.0	68.81
Amino acid	Histamine	67.99
Amino acid	L-Threonine	67.85
Amino acid	L-Glutamine	67.70
Lipid	PC (38:3)	65.30
Organic acid	Pyruvic.acid	64.96

The most important metabolites in demarcating the ApoE4AD phenotypes were putrescine, Spha.1.P-(18:0), uracil, triglyceride (56:0), 3- hydroxybutyric acid, sarcosine, phosphoethanolamine (38:5), LPA (20:5), L-tryptophan, sphingomyelin (d18:1/20:1), glutathione, L-serine, L-histidine, LPA (18:0), histamine, L-threonine, L-glutamine, phosphatidylcholine (38:3), see Table 6.

**3.3. Metabolite Covariance Network Analysis.** The Gaussian graphical modelling between ApoE4 carriers and non-carriers with AD showed distinct correlations between the metabolites (Fig. 11). The ApoE4 carriers' network was less sparse, having less edges compared to the non-carriers'. Moreover, the communities of covarying metabolites are distinct among the two groups, as shown in Fig. 12.

The Wilcoxon Signed Rank Test showed higher degrees of centrality in ApoE4 carriers compared to non-carriers ( $p = 0.001$ ). In the ApoE4 positive group, central metabolites were 1-methylhistidine, citrulline and pyroglutamic acid, followed by asymmetric dimethylarginine (ADMA), symmetric dimethylarginine (SDMA), 2-ketoglutaric acid, uracil, histamine, L-arginine, 5 organic acids, 3 lipids and 2 oxidative stress compounds (Table 7). ApoE4 non-carriers had 3-methylhistidine as top central metabolite, followed by lyso-sphingolipid 1-P(18:1), methoxytyrosine, 3-aminoisobutyric acid, dopamine, putrescine, glutamate and succinate, sphingomyelin (d18:1/21:0) and oxidative stress compound cLPA (18:2).

TABLE 7. Central (centrality degreee >5) metabolites in ApoE4 carriers and non-carriers with AD.

Metabolite	ApoE4 carriers		ApoE4 non-carriers	
	Metabolite	Degree	Metabolite	Degree
Am.X1.Methylhistidine	7	Am.X3.Methylhistidine	7	
Am.Citrulline	7	Lip.SM.d18.1.20.0.	6	
OA.Pyroglutamic.acid	7	Am.X3.Methoxytyrosine	5	
Am.ADMA	6	Am.DL.3.aminoisobutyric.acid	5	
Am.SDMA	6	Am.Dopamine	5	
OA.2.ketoglutaric.acid	6	Am.Putrescine	5	
OA.Uracil	6	OA.Glutamic.Acid	5	
Am.Histamine	5	OA.Succinic.acid	5	
Am.L.Arginine	5	Lip.SM.d18.1.21.0.	5	
Am.L.Kynurenone	5	OS.HpH.cLPA.C18.2	5	
Am.N6.N6.N6.Trimethyl.L.lysine	5			
Am.Putrescine	5			
OA.Lactic.acid	5			
OA.Fumaric.acid	5			
OA.Aspartic.acid	5			
OA.Iminodiacetate	5			
OA.3.hydroxyisovaleric.acid	5			
Lip.PC.36.5.	5			
Lip.SM.d18.1.21.0.	5			
Lip.SM.d18.1.23.0.	5			
OS.HpH.cLPA.C18.2	5			
OS.cHpH.cLPA.C20.3	5			

ADMA: Asymmetric Dimethylarginine, SD, Am: Aminoacid, Lip: Lipid, DG: Diglyceride, LPA: Lyso-sphingolipid , SM: Sphingomyelin, TG: Triglyceride, OA: Organic Acid, OS: Oxidative Stress compound, PC: Phosphatidylcholine, SDMA: Symmetric Dimethylarginine

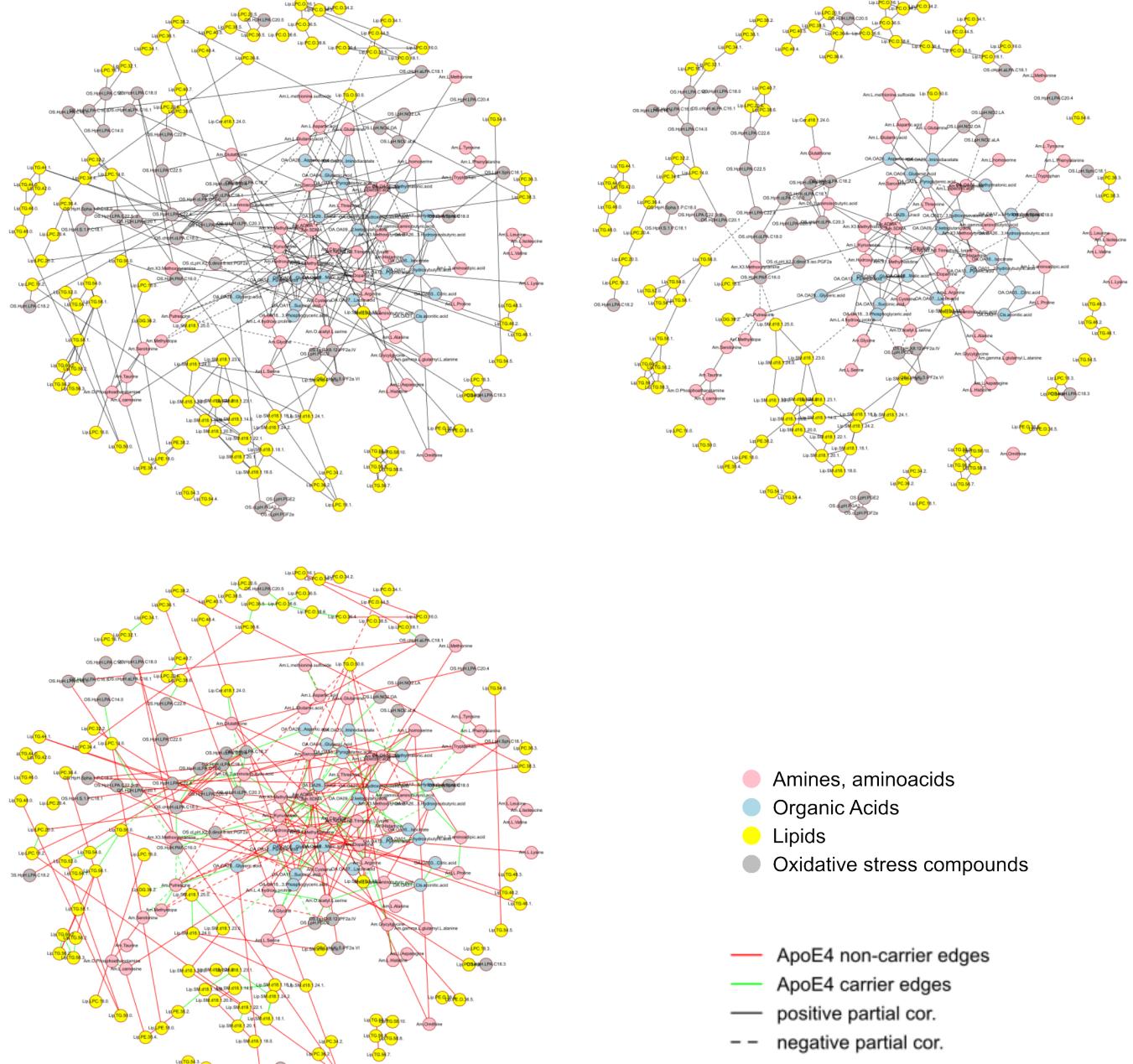


FIGURE 11. Metabolite covariance network topologies among ApoE4 non-carriers (top-left), carriers (top-right) and differential edge network (bottom) in AD.

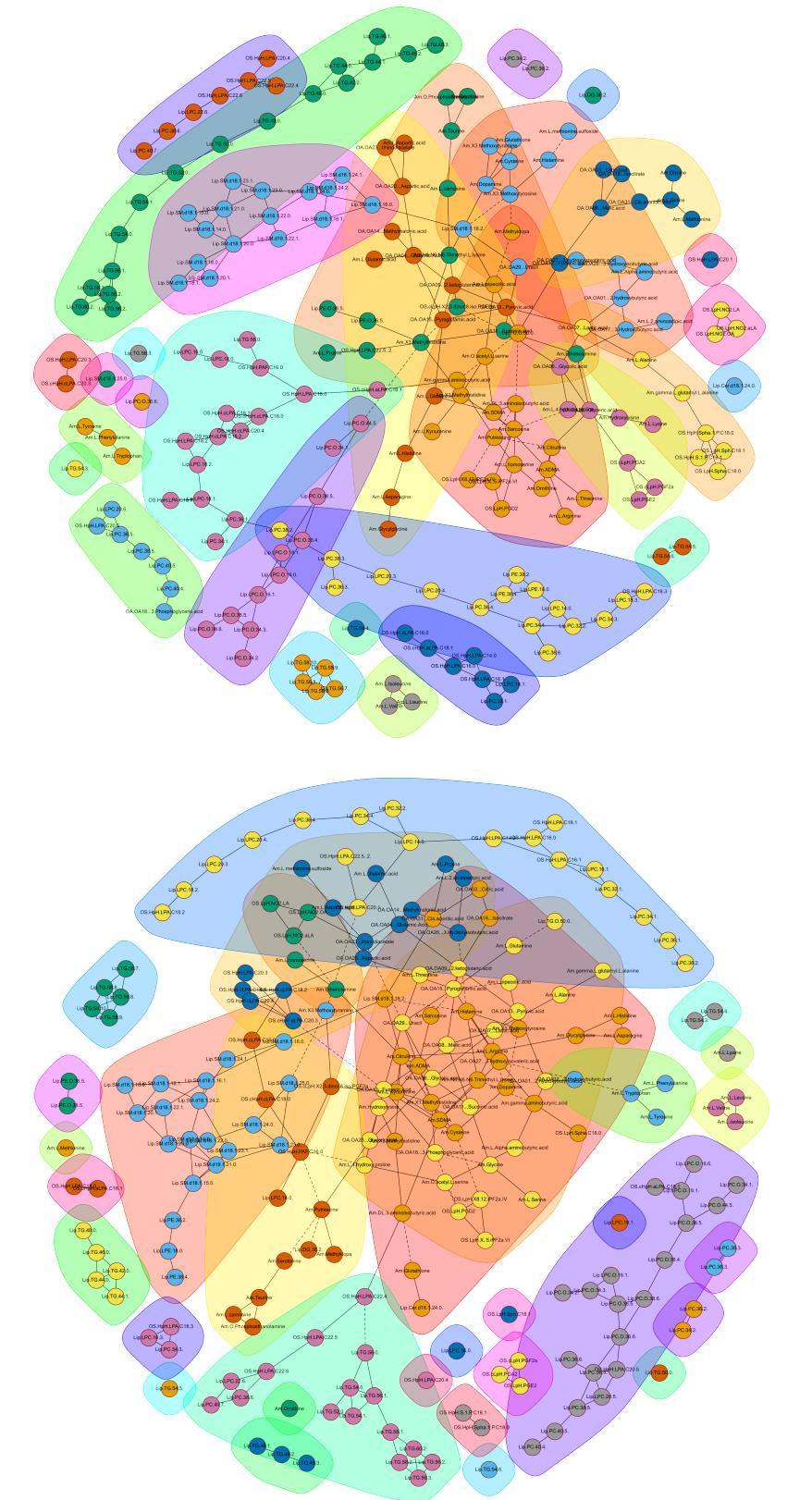


FIGURE 12. Communities of (densely correlated) metabolites in ApoE4 non-carriers (top) and ApoE4 carriers with AD (bottom).

#### 4. DISCUSSION

This study had two objectives: to unveil potentially unknown pathways of systemic metabolic deregulation attributed to ApoE4 (dose or presence), and to propose a comprehensive methodology for high-dimensional data analysis in the context of AD. To this end, serum metabolic signatures were assessed among SCD and AD with three approaches: differential expression, multiclass classification and network analysis. Differential expression highlights metabolites whose levels are shifted by ApoE4 dose. The multiclass classification reveals metabolites that, on top of clinical background variables, can be fitted to predict AD and ApoE4 status. Covariance network analysis reveals metabolic interdependencies among ApoE4 carriers and non-carriers in AD. The serum metabolites reported via these metabolic signatures here might serve as a *proxy* of metabolic perturbations in the brain –  were measured in serum obtained from peripheral blood, an (almost) non-invasive alternative to spinal aspiration for CSF.

The metabolites that emerged from the analysis were generally different between the approaches, and among AD-SCD (it is  important to note that the SCD group is not a control group, as subjects with varying degrees of cognitive impairment were included, who might develop AD or another form of dementia later). Nevertheless, certain metabolites appeared in all, or many of the metabolic signatures.

A metabolite that appeared in all three metabolic signatures was putrescine: trends of positive ApoE4 dose effects in AD, top metabolite in demarcating ApoE4 and AD status and central metabolite (5 dependent metabolites regardless of ApoE4 status) in AD. Putrescine appeared increased in a recent study [111], where it is described as a toxic by-product of activated urea cycle in  $\text{A}\beta$ -reactive astrocytes of AD brains [111, 112].

Trends of a negative  ApoE4 dose effect on putrescine and four amino acids were observed in the SCD group: L-serine, L-tryptophan, glycine and L-homoserine. Tryptophan and serine proved also important in demarcating ApoE4 and AD status. In the brain, L-tryptophan is catabolised to kynurenine with the enzymes indoleamine 2,3-dioxygenase and tryptophan 2,3-dioxygenase, through the kynurenine pathway [113]. In this study, kynurenine was found central in the serum metabolome, branched with five metabolites. Recent reviews  kynurenine pathway metabolites closely associated with AD pathogenesis [113, 114]. Increased kynurenine pathway metabolites may reflect the extent of neuroinflammation in amyotrophic lateral sclerosis, frontotemporal dementia and early onset AD [115]. A whole-blood targeted metabolomics panel also revealed increased kynurenine pathway metabolites [116]. Conversely, L-serine levels appear to be deficient in AD brains due to impaired glycolysis [117].

ApoE4 showed trends of a negative  effect on ketoglutarate in the global test and the ANCOVA F-tests exclusively in the AD group. The network analysis revealed it is also a central metabolite, driving the variance of six other metabolites. In the brain, the  $\alpha$ -ketoglutarate dehydrogenase complex is perceived as a "hub of plasticity in neurodegeneration and regeneration", as it reflects the impaired glucose metabolism and key enzymes of the tricarboxylic acid cycle [118].

**ApoE4 dose effects on serum metabolite levels** The global test for ApoE4 dose effects on serum metabolite levels revealed positive dose effects on lipids in AD, mainly triglycerides and diglycerides. This is consistent with evidence from Caucasian [119–121] and southern Chinese populations [122] that shows elevated TGs associated with  ApoE4. The same test was not significant in the SCD group; mostly amines and amino acids showed trends of an effect. This could putatively reflect compensatory mechanisms still in place for the debilitating effects of ApoE4.

The nested linear model comparison did not reveal any significant ApoE4 dose effects, after correcting FDR. This can be partially explained considering the covariates in each approach. The global test was corrected for sex only, while the nested models fitted several background clinical factors. That is, the metabolic variance explained by ApoE4 dose in the global test might be partially attributed to other clinical factors in the nested models, such as hypercholesterolemia (that often coexists  with hypertriglyceridemia and is not measured in this study). Notably, the metabolites showing trends of  an Apoe4 dose effect were

different between the AD and SCD group. Consistent with the global test, AD group presented mainly a lipid (TG & DG) signature, while the SCD group an amino one (serine, tryptophan, glycine, homoserine and putrescine).

**Classification of ApoE4 and AD status** Serum metabolic information, in the form of a semi-targeted metabolomic panel or its latent 6-factor projection seems to slightly increase the classification potential of the models. Interestingly, all models discriminated better among certain classes (AD+E4 vs. SCD+E4, AD+E4 vs. SCD, AD-E4 vs. SCD+E4 and AD-E4 vs. SCD-E4) compared to others (AD+E4 vs. AD-F<sup>1</sup> and SCD+E4 vs. SCD-E4). This can be putatively attributed to the class imbalance between AD+E4 (3 ), and AD () (86), as well as SCD+E4 () and SCD () (40). It might also show that the metabolic nuances among ApoE4 carriers and non-carriers are not as pronounced as among AD and SCD.

Each of the top four metabolites  demarcating  E4 and AD status belonged to one of the major metabolite classes studied: amines (putrescine), oxidative stress compounds (sphingolipid 1.P.C18.0), organic acids (uracil) and lipids (TG (56:0)).

### Network analysis

## 5. CONCLUSION

## REFERENCES

- Penke, B., Szűcs, M. & Bogár, F. New Pathways Identify Novel Drug Targets for the Prevention and Treatment of Alzheimer's Disease. *International Journal of Molecular Sciences* **24**, 5383. issn: 1422-0067. <https://www.mdpi.com/1422-0067/24/6/5383> (Mar. 2023).
- 2023 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **19**, 1598–1695. issn: 1552-5279. <https://onlinelibrary.wiley.com/doi/full/10.1002/alz.13016%20https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.13016%20https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13016> (Apr. 2023).
- Heneka, M. T. et al. Neuroinflammation in Alzheimer's disease. *The Lancet. Neurology* **14**, 388–405. issn: 1474-4465. <https://pubmed.ncbi.nlm.nih.gov/25792098/> (Apr. 2015).
- Edwards, F. A. A Unifying Hypothesis for Alzheimer's Disease: From Plaques to Neurodegeneration. *Trends in Neurosciences* **42**, 310–322. issn: 1878108X. [http://www.cell.com/article/S016622361930027X/fulltext%20http://www.cell.com/article/S016622361930027X/abstract%20https://www.cell.com/trends/neurosciences/abstract/S0166-2236\(19\)30027-X](http://www.cell.com/article/S016622361930027X/fulltext%20http://www.cell.com/article/S016622361930027X/abstract%20https://www.cell.com/trends/neurosciences/abstract/S0166-2236(19)30027-X) (May 2019).
- Aaldijk, E. & Vermeiren, Y. The role of serotonin within the microbiota-gut-brain axis in the development of Alzheimer's disease: A narrative review. *Ageing Research Reviews* **75**, 101556. issn: 1568-1637. <https://www.sciencedirect.com/science/article/pii/S1568163721003032> (2022).
- Scheltens, P. et al. Alzheimer's disease. *Lancet (London, England)* **388**, 505–517. issn: 1474-547X. <https://pubmed.ncbi.nlm.nih.gov/26921134/> (July 2016).
- Beydoun, M. A. et al. Epidemiologic studies of modifiable factors associated with cognition and dementia: Systematic review and meta-analysis. *BMC Public Health* **14**, 1–33. issn: 14712458. <https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-14-643> (June 2014).
- Van Cauwenbergh, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genetics in Medicine* **2016** *18*:5 **18**, 421–430. issn: 1530-0366. <https://www.nature.com/articles/gim2015117> (Aug. 2015).
- Hardy, J. Alzheimer's disease: The amyloid cascade hypothesis: An update and reappraisal. *Journal of Alzheimer's Disease* **9**, 151–153. issn: 1387-2877 (Jan. 2006).
- Kepp, K. P., Robakis, N. K., Høilund-Carlsen, P. F., Sensi, S. L. & Vissel, B. The amyloid cascade hypothesis: an updated critical review. <https://doi.org/10.1093/brain/awad159> (2023).
- Kurkinen, M. et al. The Amyloid Cascade Hypothesis in Alzheimer's Disease: Should We Change Our Thinking? *Biomolecules* **2023**, Vol. 13, Page 453 **13**, 453. issn: 2218-273X. <https://www.mdpi.com/2218-273X/13/3/453> (Mar. 2023).
- Corder, E. H. et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923. issn: 00368075 (Aug. 1993).
- Husain, M. A., Laurent, B. & Plourde, M. APOE and Alzheimer's Disease: From Lipid Transport to Physiopathology and Therapeutics. *Frontiers in Neuroscience* **15**, 630502. issn: 1662453X (Feb. 2021).
- Yang, L. G., March, Z. M., Stephenson, R. A. & Narayan, P. S. Apolipoprotein E in lipid metabolism and neurodegenerative disease. *Trends in Endocrinology & Metabolism* **34**, 430–445. issn: 1043-2760. [http://www.cell.com/article/S1043276023000929/fulltext%20http://www.cell.com/article/S1043276023000929/abstract%20https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760\(23\)00092-9](http://www.cell.com/article/S1043276023000929/fulltext%20http://www.cell.com/article/S1043276023000929/abstract%20https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760(23)00092-9) (Aug. 2023).
- Liu, C. & al. et, e. Apolipoprotein E and Alzheimer disease: risk, mechanisms, and therapy. *Nat. Rev. Neurosci.* **9**, 106–118 (2013).
- Strittmatter, W. & al. et, e. Apolipoprotein E: high-avidity binding to  $\beta$ -amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1977–1981 (1993).
- Deming, Y. & al. et, e. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol.* **133**, 839–856 (2017).
- Belloy, M. & al. et, e. A quarter century of APOE and Alzheimer's disease: progress to date and the path forward. *Neuron* **101**, 820–838 (2019).
- Farrer, L. & al. et, e. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *JAMA* **278**, 1349–1356 (1997).
- Bookheimer, S. Y. & Burggren, A. C. APOE-4 genotype and neurophysiological vulnerability to Alzheimer's and cognitive aging. *Annual review of clinical psychology* **5**, 343–62. <https://api.semanticscholar.org/CorpusID:207677324> (2009).
- Kim, J., Basak, J. M. & Holtzman, D. M. The Role of Apolipoprotein E in Alzheimer's Disease. *Neuron* **63**, 287–303. <https://api.semanticscholar.org/CorpusID:40937173> (2009).
- Serrano-Pozo, A. & Growdon, J. Is Alzheimer's Disease Risk Modifiable? *Journal of Alzheimer's disease : JAD* **67** *3*, 795–819. <https://api.semanticscholar.org/CorpusID:73482621> (2019).
- Arnold, M. et al. Sex and APOE  $\epsilon 4$  genotype modify the Alzheimer's disease serum metabolome. *Nature Communications* **2020** *11*:1 **11**, 1–12. issn: 2041-1723. <https://www.nature.com/articles/s41467-020-14959-w> (Mar. 2020).

24. Yassine, H. N. & Finch, C. E. APOE Alleles and Diet in Brain Aging and Alzheimer's Disease. *Frontiers in Aging Neuroscience* **12**, 544681. issn: 16634365 (June 2020).
25. Egert, S., Rimbach, G. & Huebbe, P. ApoE genotype: from geographic distribution to function and responsiveness to dietary factors. *Proceedings of the Nutrition Society* **71**, 410–424. issn: 1475-2719. <https://www.cambridge.org/core/journals/proceedings-of-the-nutrition-society/article/apoe-genotype-from-geographic-distribution-to-function-and-responsiveness-to-dietary-factors/D9F35B3FFE46F48CAFE131E0E94CC62C> (Aug. 2012).
26. Eisenberg, D. T., Kuzawa, C. W. & Hayes, M. G. Worldwide allele frequencies of the human apolipoprotein E gene: Climate, local adaptations, and evolutionary history. *American Journal of Physical Anthropology* **143**, 100–111. issn: 1096-8644. <https://onlinelibrary.wiley.com/doi/full/10.1002/ajpa.21298%20https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.21298> (Sept. 2010).
27. Logue, M. W. *et al.* A Comprehensive Genetic Association Study of Alzheimer Disease in African Americans. *Archives of Neurology* **68**, 1569–1579. issn: 0003-9942. <https://jamanetwork.com/journals/jamaneurology/fullarticle/1108047> (Dec. 2011).
28. Blue, E. E., Horimoto, A. R., Mukherjee, S., Wijsman, E. M. & Thornton, T. A. Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. *Alzheimer's & Dementia* **15**, 1524–1532. issn: 1552-5279. [https://onlinelibrary.wiley.com/doi/full/10.1016/j.jalz.2019.07.016%20https://onlinelibrary.wiley.com/doi/abs/10.1016/j.jalz.2019.07.016](https://onlinelibrary.wiley.com/doi/full/10.1016/j.jalz.2019.07.016%20https://onlinelibrary.wiley.com/doi/abs/10.1016/j.jalz.2019.07.016%20https://alz-journals.onlinelibrary.wiley.com/doi/10.1016/j.jalz.2019.07.016) (Dec. 2019).
29. Suchy-Dicey, A., Howard, B., Longstreth, W. T., Reiman, E. M. & Buchwald, D. APOE genotype, hippocampus, and cognitive markers of Alzheimer's disease in American Indians: Data from the Strong Heart Study. *Alzheimer's & Dementia* **18**, 2518–2526. issn: 1552-5279. [https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12573%20https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.12573](https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12573%20https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.12573%20https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12573) (Dec. 2022).
30. Rajabli, F. *et al.* Ancestral origin of ApoE e4 Alzheimer disease risk in Puerto Rican and African American populations. *PLOS Genetics* **14**, e1007791. issn: 1553-7404. <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007791> (Dec. 2018).
31. Naslavsky, M. S. *et al.* Global and local ancestry modulate APOE association with Alzheimer's neuropathology and cognitive outcomes in an admixed sample. *Molecular Psychiatry* **2022** *27:11* **27**, 4800–4808. issn: 1476-5578. <https://www.nature.com/articles/s41380-022-01729-x> (Sept. 2022).
32. Chen, H.-K. & al. et. e. Apolipoprotein E4 domain interaction mediates detrimental effects on mitochondria and is a potential therapeutic target for Alzheimer disease. *J. Biol. Chem.* **286**, 5215–5221 (2011).
33. Rajabli, F. *et al.* A locus at 19q13.31 significantly reduces the ApoE e4 risk for Alzheimer's Disease in African Ancestry. *PLOS Genetics* **18**, e1009977. issn: 1553-7404. <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009977> (July 2022).
34. Hunsberger, H. C., Pinky, P. D., Smith, W., Suppiramaniam, V. & Reed, M. N. The role of APOE4 in Alzheimer's disease: strategies for future therapeutic interventions. *Neuronal Signaling* **3**, 1–15. issn: 20596553. <https://neuronalsignal/article/3/2/NS20180203/110989/The-role-of-APOE4-in-Alzheimer-s-disease%20https://dx.doi.org/10.1042/NS20180203> (June 2019).
35. Trumble, B. C. *et al.* Apolipoprotein E4 is associated with improved cognitive function in Amazonian forager-horticulturalists with a high parasite burden. *FASEB Journal* **31**, 1508–1515. issn: 15306860 (Apr. 2017).
36. Finch, C. E. & Sapolsky, R. M. The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms. *Neurobiology of Aging* **20**, 407–428. issn: 01974580 (1999).
37. Mahley, R. Central nervous system lipoproteins: ApoE and regulation of cholesterol metabolism. *Arterioscler. Thromb. Vasc. Biol.* **36**, 1305–1315 (2016).
38. Lanfranco, M. & al. et. e. Expression and secretion of apoE isoforms in astrocytes and microglia during inflammation. *Glia* **69**, 1478–1493 (2021).
39. Castellano, J. M. *et al.* Human apoE isoforms differentially regulate brain amyloid- $\beta$  peptide clearance. *Science Translational Medicine* **3**. issn: 19466234 (June 2011).
40. Cruchaga, C. *et al.* Cerebrospinal fluid APOE levels: An endophenotype for genetic studies for Alzheimer's disease. *Human Molecular Genetics* **21**, 4558–4571. issn: 09646906 (Oct. 2012).
41. Bu, G. APOE targeting strategy in Alzheimer's disease: lessons learned from protective variants. *Molecular Neurodegeneration* **17**, 1–4. issn: 17501326. <https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-022-00556-6> (Dec. 2022).
42. Hu, J. *et al.* Opposing effects of viral mediated brain expression of apolipoprotein E2 (apoE2) and apoE4 on apoE lipidation and A $\beta$  metabolism in apoE4-targeted replacement mice. *Molecular Neurodegeneration* **10**, 1–11. issn: 17501326. <https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-015-0001-3> (Mar. 2015).
43. Flowers, S. A. & Rebeck, G. W. APOE in the normal brain. *Neurobiology of Disease* **136**, 104724. issn: 0969-9961 (Mar. 2020).

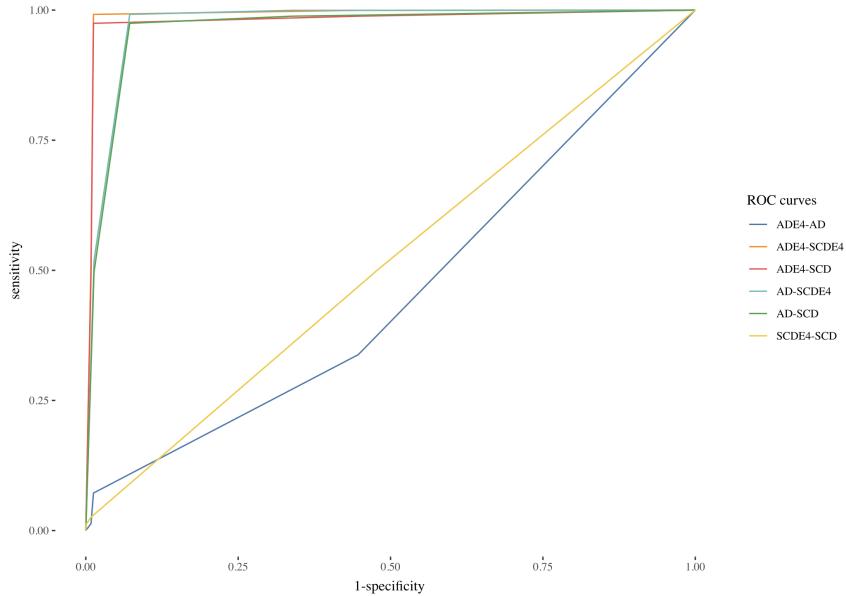
44. Courtney, R. & Landreth, G. E. LXR Regulation of Brain Cholesterol: From Development to Disease. *Trends in Endocrinology and Metabolism* **27**, 404–414. ISSN: 18793061. [http://www.cell.com/article/S1043276016300042/fulltext%20http://www.cell.com/article/S1043276016300042/abstract%20https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760\(16\)30004-2](http://www.cell.com/article/S1043276016300042/fulltext%20http://www.cell.com/article/S1043276016300042/abstract%20https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760(16)30004-2) (June 2016).
45. Heinsinger, N. M., Gachechiladze, M. A. & Rebeck, G. W. Apolipoprotein E Genotype Affects Size of ApoE Complexes in Cerebrospinal Fluid. *Journal of Neuropathology & Experimental Neurology* **75**, 918–924. ISSN: 0022-3069. <https://doi.org/10.1093/jnen/nlw067> (Oct. 2016).
46. Nguyen, D. *et al.* Molecular basis for the differences in lipid and lipoprotein binding properties of human apolipoproteins E3 and E4. *Biochemistry* **49**, 10881–10889. ISSN: 00062960. <https://pubs.acs.org/doi/abs/10.1021/bi1017655> (Dec. 2010).
47. Hatters, D. M., Peters-Libeu, C. A. & Weisgraber, K. H. Apolipoprotein E structure: insights into function. *Trends in Biochemical Sciences* **31**, 445–454. ISSN: 09680004 (Aug. 2006).
48. Hatters, D. M., Zhong, N., Rutener, E. & Weisgraber, K. H. Amino-terminal Domain Stability Mediates Apolipoprotein E Aggregation into Neurotoxic Fibrils. *Journal of Molecular Biology* **361**, 932–944. ISSN: 00222836 (Sept. 2006).
49. Rawat, V. *et al.* ApoE4 Alters ABCA1 Membrane Trafficking in Astrocytes. *Journal of Neuroscience* **39**, 9611–9622. ISSN: 0270-6474. <https://www.jneurosci.org/content/39/48/9611%20https://www.jneurosci.org/content/39/48/9611.abstract> (Nov. 2019).
50. Zhao, N. *et al.* Apolipoprotein E4 Impairs Neuronal Insulin Signaling by Trapping Insulin Receptor in the Endosomes. *Neuron* **96**, 115–129. ISSN: 1097-4199. <https://pubmed.ncbi.nlm.nih.gov/28957663/> (Sept. 2017).
51. Svennerholm, L., Boström, K. & Jungbjer, B. Changes in weight and compositions of major membrane components of human brain during the span of adult human life of Swedes. *Acta Neuropathologica* **94**, 345–352. ISSN: 00016322. <https://link.springer.com/article/10.1007/s004010050717> (Oct. 1997).
52. Marottoli, F. M. *et al.* Peripheral inflammation, apolipoprotein E4, and amyloid- $\beta$  interact to induce cognitive and cerebrovascular dysfunction. *ASN Neuro* **9**. ISSN: 17590914. <https://journals.sagepub.com/doi/10.1177/1759091417719201> (Aug. 2017).
53. Teng, Z. *et al.* ApoE Influences the Blood-Brain Barrier Through the NF- $\kappa$ B/MMP-9 Pathway After Traumatic Brain Injury. *Scientific Reports* **2017** *7:1* 7, 1–8. ISSN: 2045-2322. <https://www.nature.com/articles/s41598-017-06932-3> (July 2017).
54. Kloske, C. M. & Wilcock, D. M. The Important Interface Between Apolipoprotein E and Neuroinflammation in Alzheimer's Disease. *Frontiers in Immunology* **11**. ISSN: 16643224 (Apr. 2020).
55. Huang, Y. W. A., Zhou, B., Wernig, M. & Südhof, T. C. ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and  $A\beta$  Secretion. *Cell* **168**, 427–441. ISSN: 10974172 (Jan. 2017).
56. Tachibana, M. *et al.* Rescuing effects of RXR agonist bexarotene on aging-related synapse loss depend on neuronal LRP1. *Experimental Neurology* **277**, 1–9. ISSN: 0014-4886 (Mar. 2016).
57. Safieh, M., Korczyn, A. D. & Michaelson, D. M. ApoE4: an emerging therapeutic target for Alzheimer's disease. *BMC Medicine* **2019** *17:1* 17, 1–17. ISSN: 1741-7015. <https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1299-4> (Mar. 2019).
58. Cao, J. *et al.* ApoE4-associated phospholipid dysregulation contributes to development of Tau hyper-phosphorylation after traumatic brain injury. *Scientific Reports* **7**. ISSN: 20452322 (Dec. 2017).
59. Shi, Y. *et al.* ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature* **2017** *549*:7673 **549**, 523–527. ISSN: 1476-4687. <https://www.nature.com/articles/nature24016> (Sept. 2017).
60. Vasilevskaya, A. *et al.* Interaction of APOE4 alleles and PET tau imaging in former contact sport athletes. *NeuroImage: Clinical* **26**, 102212. ISSN: 2213-1582 (Jan. 2020).
61. Wang, C. & al. et. e. Gain of toxic apolipoprotein E4 effects in human iPSC-derived neurons is ameliorated by a small-molecule structure corrector. *Nat. Med.* **24**, 647–657 (2018).
62. Hoe, H. S., Freeman, J. & Rebeck, G. W. Apolipoprotein e decreases tau kinases and phospho-tau levels in primary neurons. *Molecular Neurodegeneration* **1**. ISSN: 17501326 (2006).
63. Barupal, D. K. *et al.* Sets of coregulated serum lipids are associated with Alzheimer's disease pathophysiology. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **11**, 619–627. ISSN: 2352-8729 (Dec. 2019).
64. Fernández-Calle, R. *et al.* APOE in the bullseye of neurodegenerative diseases: impact of the APOE genotype in Alzheimer's disease pathology and brain diseases. *Molecular Neurodegeneration* **2022** *17:1* 17, 1–47. ISSN: 1750-1326. <https://link.springer.com/articles/10.1186/s13024-022-00566-4%20https://link.springer.com/article/10.1186/s13024-022-00566-4> (Sept. 2022).
65. Proitsi, P. *et al.* Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. *Alzheimer's & Dementia* **13**, 140–151. ISSN: 1552-5260 (Feb. 2017).
66. Oeckl, P. *et al.* Glial Fibrillary Acidic Protein in Serum is Increased in Alzheimer's Disease and Correlates with Cognitive Impairment. *Journal of Alzheimer's disease : JAD* **67**, 481–488. ISSN: 1875-8908. <https://pubmed.ncbi.nlm.nih.gov/30594925/> (2019).

67. Bandaru, V. V. R. *et al.* ApoE4 disrupts sterol and sphingolipid metabolism in Alzheimer's but not normal brain. *Neurobiology of Aging* **30**, 591–599. issn: 0197-4580 (Apr. 2009).
68. Novotny, B. C. *et al.* Metabolomic and lipidomic signatures in autosomal dominant and late-onset Alzheimer's disease brains. *Alzheimer's & Dementia* **19**, 1785–1799. issn: 1552-5279. <https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12800><https://onlinelibrary.wiley.com/doi/10.1002/alz.12800> (May 2023).
69. Area-Gomez, E. *et al.* APOE4 is Associated with Differential Regional Vulnerability to Bioenergetic Deficits in Aged APOE Mice. *Scientific Reports* **2020 10:1** **10**, 1–20. issn: 2045-2322. <https://www.nature.com/articles/s41598-020-61142-8> (Mar. 2020).
70. Zhao, N. *et al.* Alzheimer's Risk Factors Age, APOE Genotype, and Sex Drive Distinct Molecular Pathways. *Neuron* **106**, 727–742. issn: 10974199. [http://www.cell.com/article/S0896627320301859/fulltext%20http://www.cell.com/article/S0896627320301859/abstract%20https://www.cell.com/neuron/abstract/S0896-6273\(20\)30185-9](http://www.cell.com/article/S0896627320301859/fulltext%20http://www.cell.com/article/S0896627320301859/abstract%20https://www.cell.com/neuron/abstract/S0896-6273(20)30185-9) (June 2020).
71. Peña-bautista carmen, c. *et al.* Metabolomics study to identify plasma biomarkers in alzheimer disease: ApoE genotype effect. *Journal of Pharmaceutical and Biomedical Analysis* **180**, 113088. issn: 0731-7085 (Feb. 2020).
72. Farmer, B. & al. et, e. APOE4 lowers energy expenditure in females and impairs glucose oxidation by increasing flux through aerobic glycolysis. *Mol. Neurodegener.* **16**, 62 (2021).
73. De Leeuw, F. A. *et al.* Blood-based metabolic signatures in Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **8**, 196–207. issn: 2352-8729 (Jan. 2017).
74. Green, R. E. *et al.* Investigating associations between blood metabolites, later life brain imaging measures, and genetic risk for Alzheimer's disease. *Alzheimer's Research and Therapy* **15**, 1–13. issn: 17589193. <https://alzres.biomedcentral.com/articles/10.1186/s13195-023-01184-y%20http://creativecommons.org/publicdomain/zero/1.0/> (Dec. 2023).
75. Varma, V. R. *et al.* Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. *PLOS Medicine* **15**, e1002482. issn: 1549-1676. <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002482> (Jan. 2018).
76. Sun, L. *et al.* Association between Human Blood Metabolome and the Risk of Alzheimer's Disease. *Annals of Neurology* **92**, 756–767. issn: 1531-8249. <https://onlinelibrary.wiley.com/doi/full/10.1002/ana.26464%20https://onlinelibrary.wiley.com/doi/abs/10.1002/ana.26464%20https://onlinelibrary.wiley.com/doi/10.1002/ana.26464> (Nov. 2022).
77. Jia, L. *et al.* A metabolite panel that differentiates Alzheimer's disease from other dementia types. *Alzheimer's & Dementia* **18**, 1345–1356. issn: 1552-5279. <https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12484> (July 2022).
78. Huo, Z. *et al.* Brain and blood metabolome for Alzheimer's dementia: findings from a targeted metabolomics analysis. *Neurobiology of Aging* **86**, 123–133. issn: 0197-4580 (Feb. 2020).
79. Weng, W. C., Huang, W. Y., Tang, H. Y., Cheng, M. L. & Chen, K. H. The Differences of Serum Metabolites Between Patients With Early-Stage Alzheimer's Disease and Mild Cognitive Impairment. *Frontiers in Neurology* **10**, 482026. issn: 16642295 (Nov. 2019).
80. Simpson, B. N. *et al.* Blood metabolite markers of cognitive performance and brain function in aging. *Journal of Cerebral Blood Flow and Metabolism* **36**, 1212–1223. issn: 15597016. <https://journals.sagepub.com/doi/full/10.1177/0271678X15611678> (July 2016).
81. Oka, T., Matsuzawa, Y., Tsuneyoshi, M. & Nakamura, Y. Multiomics analysis to explore blood metabolite biomarkers in an Alzheimer's Disease Neuroimaging Initiative cohort. <https://www.researchsquare.com%20https://www.researchsquare.com/article/rs-2973576/v1> (June 2023).
82. Peeters, C. F. W. *et al.* Stable prediction with radiomics data. <https://arxiv.org/abs/1903.11696v1> (Mar. 2019).
83. Peeters, C. F., Bilgrau, A. E. & van Wieringen, W. N. rags2ridges: A One-Stop-ℓ2-Shop for Graphical Modeling of High-Dimensional Precision Matrices. *Journal of Statistical Software* **102**, 1–32. issn: 1548-7660. <https://www.jstatsoft.org/index.php/jss/article/view/v102i04> (May 2022).
84. Van Der Flier, W. M. & Scheltens, P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. *Journal of Alzheimer's disease : JAD* **62**, 1091–1111. issn: 1875-8908. <https://pubmed.ncbi.nlm.nih.gov/29562540/> (2018).
85. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **2016 3:1** **3**, 1–9. issn: 2052-4463. <https://www.nature.com/articles/sdata201618> (Mar. 2016).
86. Simon, R. M. *et al.* Design and Analysis of DNA Microarray Investigations Series: Statistics for Biology and Health, 199 (2004).
87. Goeman, J. J., Van de Geer, S., De Kort, F. & van Houwelingen, H. C. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics (Oxford, England)* **20**, 93–99. issn: 1367-4803. <https://pubmed.ncbi.nlm.nih.gov/14693814/> (Jan. 2004).
88. Goeman, J. J., Van De Geer, S. A. & Van Houwelingen, H. C. Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 477–493. issn: 1467-9868. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9868.2006.00551.x> (June 2006).

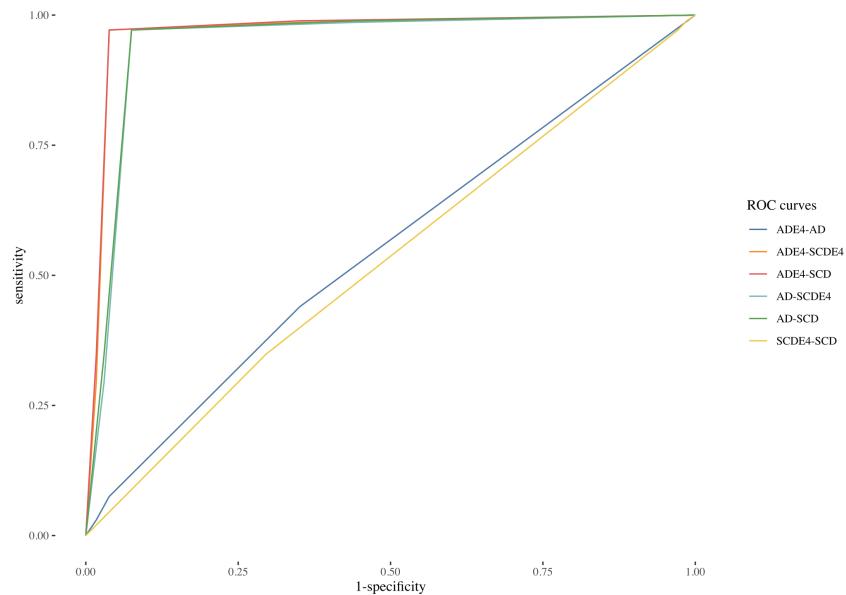
89. Goeman, J., Oosting, J., Finos, L., Solari, A. & Edelmann, D. The Global Test and the globaltest R package (2023).
90. Ott, R. L. & Longnecker, M. T. *An introduction to statistical methods and data analysis* (Cengage Learning, 2015).
91. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300. issn: 00359246 (1995).
92. Drummond, C. in *Encyclopedia of Machine Learning* (eds Sammut, C. & Webb, G. I.) 171–171 (Springer US, Boston, MA, 2010). ISBN: 978-0-387-30164-8. [https://doi.org/10.1007/978-0-387-30164-8\\_111](https://doi.org/10.1007/978-0-387-30164-8_111).
93. Elshawi, R., Al-Mallah, M. H. & Sakr, S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Medical Informatics and Decision Making* **19**, 1–32. issn: 14726947. <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0874-0> (July 2019).
94. Geman, S., Bienenstock, E. & Doursat, R. Neural Networks and the Bias/Variance Dilemma. *Neural Computation* **4**, 1–58. issn: 0899-7667. <https://dx.doi.org/10.1162/neco.1992.4.1.1> (Jan. 1992).
95. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**, 1–26. issn: 1548-7660. <https://www.jstatsoft.org/index.php/jss/article/view/v028i05> (Nov. 2008).
96. Torgo, L. *Data Mining with R, learning with case studies, 2nd edition* <http://ltorgo.github.io/DMwR2> (Chapman and Hall/CRC, 2016).
97. James, G., Witten, D., Hastie, T. & Tibshirani, R. An Introduction to Statistical Learning with Applications in R Second Edition (2023).
98. Cessie, S. I. & Houwelingen, J. C. v. Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **41**, 191–201. issn: 1467-9876. <https://onlinelibrary.wiley.com/doi/full/10.2307/2347628%20https://onlinelibrary.wiley.com/doi/abs/10.2307/2347628%20https://rss.onlinelibrary.wiley.com/doi/10.2307/2347628> (Mar. 1992).
99. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* Fourth. ISBN 0-387-95457-0. <https://www.stats.ox.ac.uk/pub/MASS4/> (Springer, New York, 2002).
100. Song, Y. Y. & Lu, Y. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry* **27**, 130. issn: 10020829. <https://pmc.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/%20https://pmc.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/?report=abstract%20https://pmc.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/> (Apr. 2015).
101. Therneau, T. M., Atkinson, B., Ripley, B., et al. rpart: Recursive partitioning. *R package version 3* (2010).
102. Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics* **28**, 337–407. <https://doi.org/10.1214/aos/1016218223> (2000).
103. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **29**, 1189–1232. issn: 00905364. <https://www.jstor.org/stable/2699986> (2001).
104. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **13-17-August-2016**, 785–794. <https://arxiv.org/abs/1603.02754v3> (Mar. 2016).
105. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
106. Barabási, A.-L. *Network Science* <http://networksciencebook.com/> (Cambridge University Press, 2015).
107. Perez De Souza, L., Alseekh, S., Brotman, Y. & Fernie, A. R. Network-based strategies in metabolomics data analysis and interpretation: from molecular networking to biological interpretation. *Expert Review of Proteomics* **17**, 243–255. issn: 17448387 (Apr. 2020).
108. Koller, D. & Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*.
109. Amara, A. et al. Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation. *Frontiers in Molecular Biosciences* **9**, 841373. issn: 2296889X (Mar. 2022).
110. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113. <https://link.aps.org/doi/10.1103/PhysRevE.69.026113> (2 Feb. 2004).
111. Ju, Y. H. et al. Astrocytic urea cycle detoxifies  $\text{A}\beta$ -derived ammonia while impairing memory in Alzheimer's disease. *bioRxiv*. <https://api.semanticscholar.org/CorpusID:239029329> (2021).
112. Wong, W. Pathogenic putrescine. *Science signaling* **15** 751, eade8161. <https://api.semanticscholar.org/CorpusID:252221531> (2022).
113. Liang, Y. et al. Kynurenone Pathway Metabolites as Biomarkers in Alzheimer's Disease. *Disease Markers* **2022**. <https://api.semanticscholar.org/CorpusID:246155814> (2022).
114. Sharma, V., Singh, T. G., Prabhakar, N. K. & Mannan, A. V. Kynurenone Metabolism and Alzheimer's Disease: The Potential Targets and Approaches. *Neurochemical Research* **47**, 1459–1476. <https://api.semanticscholar.org/CorpusID:246650872> (2022).
115. Heylen, A. et al. Brain Kynurenone Pathway Metabolite Levels May Reflect Extent of Neuroinflammation in ALS, FTD and Early Onset AD. *Pharmaceuticals* **16**. <https://api.semanticscholar.org/CorpusID:258274092> (2023).
116. Teruya, T., Chen, Y.-J., Kondoh, H., Fukuiji, Y. & Yanagida, M. Whole-blood metabolomics of dementia patients reveal classes of disease-linked metabolites. *Proceedings of the National Academy of Sciences of the United States of America* **118**. <https://api.semanticscholar.org/CorpusID:237441858> (2021).

117. Douce, J. L. *et al.* Impairment of Glycolysis-Derived l-Serine Production in Astrocytes Contributes to Cognitive Deficits in Alzheimer's Disease. *Cell metabolism* **31** 3, 503–517.e8. <https://api.semanticscholar.org/CorpusID:212416482> (2020).
118. Hansen, G. E. & Gibson, G. E. The  $\alpha$ -Ketoglutarate Dehydrogenase Complex as a Hub of Plasticity in Neurodegeneration and Regeneration. *International Journal of Molecular Sciences* **23**. <https://api.semanticscholar.org/CorpusID:253012556> (2022).
119. Maxwell, T. J. *et al.* APOE Modulates the Correlation Between Triglycerides, Cholesterol, and CHD Through Pleiotropy, and Gene-by-Gene Interactions. *Genetics* **195**, 1397–1405. <https://api.semanticscholar.org/CorpusID:1005216> (2013).
120. Carvalho-Wells, A. L., Jackson, K. G., Lockyer, S., Lovegrove, J. A. & Minihane, A. M. APOE genotype influences triglyceride and C-reactive protein responses to altered dietary fat intake in UK adults. *The American Journal of Clinical Nutrition* **96**, 1447–1453. ISSN: 0002-9165. <https://www.sciencedirect.com/science/article/pii/S0002916523029349> (2012).
121. Bernath, M. *et al.* Serum triglycerides in Alzheimer disease. *Neurology* **94**, e2088–e2098. <https://api.semanticscholar.org/CorpusID:218480337> (2020).
122. Gan, C., Zhang, Y., Liang, F., Guo, X. & Zhong, Z. Effects of APOE gene  $\epsilon 4$  allele on serum lipid profiles and risk of cardiovascular disease and tumorigenesis in southern Chinese population. *World Journal of Surgical Oncology* **20**, 1–12. ISSN: 14777819. <https://wjso.biomedcentral.com/articles/10.1186/s12957-022-02748-2> (Dec. 2022).

## APPENDIX A. MULTICLASS ROC CURVES



**FIGURE 13.** ROC curves of benchmark model: Multinomial Logistic Regression fitting the clinical background features only, obtained from repeated (100 times) 10-fold CV.



**FIGURE 14.** ROC curves of Decision Tree fitting the 6 ML-estimated latent factors on top of the clinical background variables, obtained from repeated (100 times) 10-fold CV

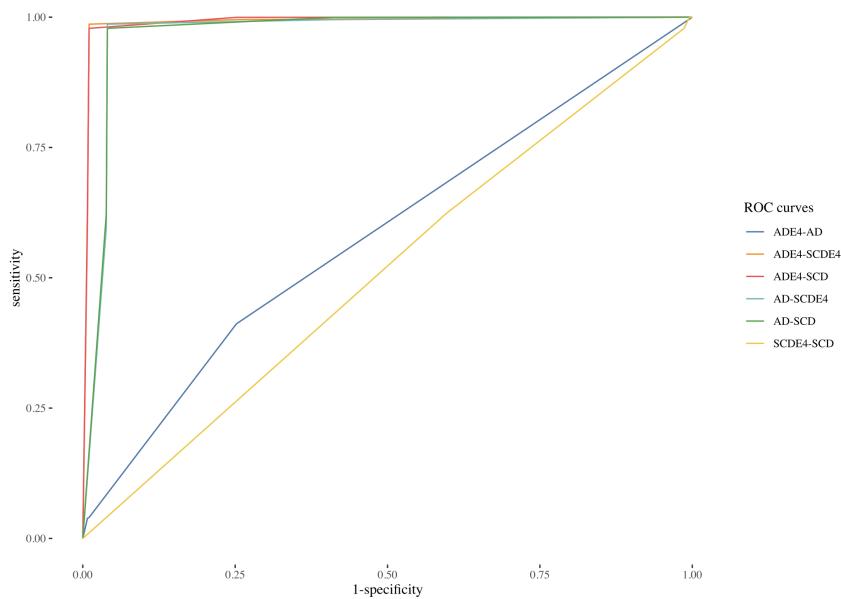


FIGURE 15. ROC curves of Xtreme Gradient Booster fitting the 6 ML-estimated latent factors on top of the clinical background variables, obtained from repeated (100 times) 10-fold CV

## APPENDIX B. R SESSION INFORMATION

```
R version 4.3.2 (2023-10-31)
Platform: aarch64-apple-darwin20 (64-bit)
Running under: macOS Sonoma 14.2.1

Matrix products: default
BLAS:     .../vecLib.framework/Versions/A/libBLAS.dylib
LAPACK: LAPACK version 3.11.0

locale:
[1] en_US.UTF-8

time zone: Europe/Amsterdam
tzcode source: internal

attached base packages:
[1] stats      graphics   grDevices datasets  utils      methods    base

other attached packages:
[1] rags2ridges_2.2.7      nnet_7.3-19          DMwR2_0.0.2
[4] dplyr_1.1.4            caret_6.0-94         lattice_0.21-9
[7] globaltest_5.56.0       survival_3.5-7       heatmaply_1.5.0
[10] xgboost_1.7.6.1        pROC_1.18.5          FMradio_1.1.1
[13] future_1.33.0          ggthemes_5.0.0       corpcor_1.6.10
[16] viridisLite_0.4.2       plotly_4.10.3         ggplot2_3.4.4
[19] rpart_4.1.21           furrr_0.3.1          viridis_0.6.4
[22] e1071_1.7-14

loaded via a namespace (and not attached):
[1] splines_4.3.2          bitops_1.0-7          tibble_3.2.1
[4] XML_3.99-0.16          lifecycle_1.0.4       globals_0.16.2
[7] magrittr_2.0.3          Hmisc_5.1-1           rmarkdown_2.25
[10] lubridate_1.9.3         zlibbioc_1.48.0       sfsmisc_1.1-16
[13] RCurl_1.98-1.13        ipred_0.9-14          lava_1.7.3
[16] S4Vectors_0.40.2        listenv_0.9.0         gRbase_2.0.1
[19] codetools_0.2-19        tidyselect_1.2.0       TSP_1.2-4
[22] jsonlite_1.8.8         Formula_1.2-5         iterators_1.0.14
[25] snowfall_1.84-6.3       Rcpp_1.0.11           glue_1.6.2
[28] tufte_0.13              TTR_0.24.4            GenomeInfoDb_1.38.2
[31] fastmap_1.1.1           fansi_1.0.6           digest_0.6.33
[34] RSQLite_2.3.4            utf8_1.2.4             tidyR_1.3.0
[37] recipes_1.0.9           class_7.3-22          httr_1.4.7
[40] gtable_0.3.4            timeDate_4032.109     blob_1.2.4
[43] RBGL_1.78.0              GSEABase_1.64.0        scales_1.3.0
[46] knitr_1.45               rstudioapi_0.15.0      tzdb_0.4.0
[49] curl_5.2.0                proxy_0.4-27           cachem_1.0.8
[52] foreign_0.8-85          AnnotationDbi_1.64.1    pillar_1.9.0
[55] VGAM_1.1-9                 xtable_1.8-4           cluster_2.1.4
```

```
[58] cli_3.6.2
[61] plyr_1.8.9
[64] Matrix_1.6-1.1
[67] quantmod_0.4.25
[70] graph_1.80.0
[73] dendextend_1.17.1
[76] DBI_1.1.3
[79] purrr_1.0.2
[82] IRanges_2.36.0
[85] annotate_1.80.0
[88] foreach_1.5.2
[91] prodlim_2023.08.28
[94] ca_0.71.1
[97] timechange_0.2.0
[100] generics_0.1.3
[103] htmlwidgets_1.6.4
[106] registry_0.5-1
[109] Biobase_2.62.0
[112] reshape2_1.4.4
[115] zoo_1.8-12
[118] grid_4.3.2
[121] htmlTable_2.4.2
[124] crayon_1.5.2
[127] munsell_0.5.0
[130] KEGGREST_1.42.0
[133] base64enc_0.1-3

compiler_4.3.2
stringi_1.8.3
hms_1.1.3
bit_4.0.5
xts_0.13.1
backports_1.4.1
RColorBrewer_1.1-3
BiocGenerics_0.48.1
RSpectra_0.16-1
parallelly_1.36.0
tools_4.3.2
gridExtra_2.3
withr_2.5.2
R6_2.5.1
renv_1.0.3
ModelMetrics_1.2.2.2
XVector_0.42.0
png_0.1-8
checkmate_2.3.1
stringr_1.5.1
reshape_0.8.9
evaluate_0.23
future.apply_1.11.0
Biostrings_2.70.1
igraph_1.6.0
webshot_0.5.5

rlang_1.1.2
assertthat_0.2.1
bit64_4.0.5
hardhat_1.3.0
MASS_7.3-60
yaml_2.3.8
expm_0.999-8
seriation_1.5.4
GenomeInfoDbData_1.2.11
stats4_4.3.2
snow_0.4-4
xfun_0.41
BiocManager_1.30.22
colorspace_2.1-0
data.table_1.14.10
pkgconfig_2.0.3
htmltools_0.5.7
gower_1.0.1
nlme_3.1-163
parallel_4.3.2
vctrs_0.6.5
readr_2.1.4
fdrtool_1.2.17
lazyeval_0.2.2
memoise_2.0.1
```