Journal of Biomedical Informatics 149 (2024) 104569

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Identifying the joint signature of brain atrophy and gene variant scores in Alzheimer's Disease

Federica Cruciani [b,*], Antonino Aparo [a], Lorenza Brusini [b], Carlo Combi [a], Silvia F. Storti [b], Rosalba Giugno [a], Gloria Menegaz [b,1], Ilaria Boscolo Galazzo [b,1], for the Alzheimer's Disease Neuroimaging Initiative[2]

[a] *Department of Computer Science, University of Verona, Verona, Italy*
[b] *Department of Engineering for Innovation Medicine, University of Verona, Verona, Italy*

ARTICLE INFO

Keywords:
Alzheimer's disease
Imaging Genetics
MRI
SKAT
Statistical learning
Transcriptomic analysis

ABSTRACT

The joint modeling of genetic data and brain imaging information allows for determining the pathophysiological pathways of neurodegenerative diseases such as Alzheimer's disease (AD). This task has typically been approached using mass-univariate methods that rely on a complete set of Single Nucleotide Polymorphisms (SNPs) to assess their association with selected image-derived phenotypes (IDPs). However, such methods are prone to multiple comparisons bias and, most importantly, fail to account for potential cross-feature interactions, resulting in insufficient detection of significant associations. Ways to overcome these limitations while reducing the number of traits aim at conveying genetic information at the gene level and capturing the integrated genetic effects of a set of genetic variants, rather than looking at each SNP individually. Their associations with brain IDPs are still largely unexplored in the current literature, though they can uncover new potential genetic determinants for brain modulations in the AD continuum. In this work, we explored an explainable multivariate model to analyze the genetic basis of the grey matter modulations, relying on the AD Neuroimaging Initiative (ADNI) phase 3 dataset. Cortical thicknesses and subcortical volumes derived from T1-weighted Magnetic Resonance were considered to describe the imaging phenotypes. At the same time the genetic counterpart was represented by gene variant scores extracted by the Sequence Kernel Association Test (SKAT) filtering model. Moreover, transcriptomic analysis was carried on to assess the expression of the resulting genes in the main brain structures as a form of validation. Results highlighted meaningful genotype–phenotype interactionsas defined by three latent components showing a significant difference in the projection scores between patients and controls. Among the significant associations, the model highlighted EPHX1 and BCAS1 gene variant scores involved in neurodegenerative and myelination processes, hence relevant for AD. In particular, the first was associated with decreased subcortical volumes and the second with decreasedtemporal lobe thickness. Noteworthy, BCAS1 is particularly expressed in the dentate gyrus. Overall, the proposed approach allowed capturing genotype–phenotype interactions in a restricted study cohort that was confirmed by transcriptomic analysis, offering insights into the underlying mechanisms of neurodegeneration in AD in line with previous findings and suggesting new potential disease biomarkers.

1. Introduction

Imaging genetics (IG) has rapidly grown in the last decades, offering the possibility to detect associations between genotype and neuroimaging data and opening new avenues to understand the genetic impact on individual's phenotypes, traits or risk of developing a disease. Indeed, the primary aim of IG is to assess the genetic architecture of brain structure and function, providing new insights into the brain mechanisms and into their role in shaping complex neurological, psychiatric and neurodegenerative disorders such as Alzheimer's disease (AD) [1,2]. AD represents the most common cause of dementia, accounting for around

[*] Corresponding author.
E-mail address: federica.cruciani@univr.it (F. Cruciani).

[1] Gloria Menegaz and Ilaria Boscolo Galazzo equally contributed as last authors to this work.

[2] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.jbi.2023.104569
Received 22 February 2023; Received in revised form 20 November 2023; Accepted 7 December 2023
Available online 15 December 2023

60%–80% of the total cases [3]. Given the trends in population aging and growth, AD is becoming one of the most burdensome diseases with more than 150 million people expected to be living with dementia worldwide by 2050, increasingly calling for next generation approaches for early diagnosis and biomarker-guided targeted therapies [4]. In recent years, technical and biological advances have sustained a shift in how the disease is considered, with AD being now conceptualized as a biological and clinical continuum covering three well-known phases (preclinical, mild cognitive impairment (MCI), and dementia) rather than as being part of the simple succession of clinically defined entities [3,5]. While the primary pathological hallmark of AD is the accumulation of abnormal proteins (mainly amyloid-β and hyperphosphorylated tau) in the brain, leading to a progressive synaptic, neuronal and axonal damage [6,7], its etiology is complex and much remains to be elucidated. As a result, an increasing number of studies is dedicated to uncovering its biological and genetic drivers, as well as brain imaging correlates, and on shading lights on their possible interplay. On the imaging side, structural magnetic resonance imaging (sMRI) represents a key element of the diagnostic criteria for the differential diagnosis and longitudinal monitoring of patients with dementia. Several studies have consistently observed both global and local atrophic changes in AD, lying along the hippocampal pathway (entorhinal cortex, hippocampus, parahippocampal gyrus and posterior cingulate cortex) in the early stages of the disease, while atrophy in temporal, parietal and frontal neocortices emerge at later stages, being associated with neuronal loss leading to language, visuospatial and behavioral impairments [6,8,9]. On the other hand, AD has a strong genetic component with more than 40 AD-associated genes/loci that have been identified by genome-wide association studies (GWASs) and sequencing projects over the last ten years, supported by large international GWAS consortia such as the International Genomics of Alzheimer's Project (IGAP) [10,11]. Segregation analyses have linked several genes to early-onset familial cases that are often explained by rare variants with a strong effect, including APP5, PSEN1, and PSEN2 [12]. Conversely, common risk variants for the more complex late-onset type of AD have been identified thanks to the analyses of massive GWAS data, with strongest genetic risk loci represented by TOMM40, APOE, CLU, PICALM and ADAM10 among the others [10,12]. Combining the genetic information with quantitative neuroimaging traits to unravel the genetic causes of AD nicely fits within the IG framework and is increasingly pursued in recent years, as demonstrated in several reviews on the topic [1,13]. Advances in this respect have been fostered by well-know large-scale projects such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [14,15], the UK Biobank [16] and the Enhancing Neuro Imaging Genetics through Meta Analysis (ENIGMA) consortium [17]. ADNI in particular represents the landmark AD biomarker study, being a large and rich repository of open-source genomics, neuroimaging (MRI and positron emission tomography), cognitive, behavioral, and clinical data. In particular, the last phase ADNI-3, started in 2016 and still ongoing, has introduced updated MRI technologies [15] which are still partly investigated in the current literature.

As such, these initiatives have facilitated the availability of large databases coupling imaging and genetics data acquired on the same subjects, greatly promoting the development of novel methodologies and applications in IG. The earliest IG studies focused on analyzing the influence of candidate genes and/or specific genetic variants on a series of brain image-derived endophenotypes (IDPs), usually modeled as separate outcome variables in univariate or mass-univariate approaches [2]. These studies with candidate genes and candidate IDPs have proven the validity of the IG approach, allowing scientists to test biologically plausible hypotheses and to cast light on the ways in which genetic variants shape brain morphology and functionality in different disorders including AD. However, such methods do not account for potential cross-feature interactions, in particular, they might ignore the genetic correlation among multiple phenotypes (pleiotropy) and are highly prone to multiple comparison problems leading to underpowered discoveries of significant associations [1,18,19]. Of note, multiple comparisons relate to artificially increase the likelihood of obtaining significant results by chance alone when conducting numerous statistical tests on the same data, leading to an inflated overall significance [20] which needs for stringent *post-hoc* corrections, directly related to the number of statistical tests, possibly hiding significant associations. Moreover, focusing on a single variable at a time might misattribute the nature of genetic effects on the brain and bias the interpretation of the results considering the complex relationships between genetics and IDPs, especially when effects are spatially distributed and encompass the whole brain [21].

As such, multivariate analysis methods are being increasingly exploited in this domain, in order to improve the discovery of multiple genotype–phenotype associations while circumventing the limitations inherent to univariate approaches. Methods able to capture the integrated genetic effects of a set of genetic variants rather than considering each single SNP might help performing whole-brain association studies, for example relying on polygenic risk scores (PRSs) [22] or SNP set approaches. For the latter, recent strategies [23–25] have proposed grouping SNPs together into SNP sets and testing their association with diseases instead of using individual SNPs. Common grouping strategies involve aggregating SNPs based on their location in a gene, haplotype blocks given by linkage disequilibrium (LD) or according to a given pathway. The Sequence Kernel Association Test (SKAT) represents in particular one of the most widely used SNP set approaches, being a flexible and computationally efficient logistic kernel-machine regression method to test for association between genetic variants in a region and a given trait while adjusting for covariates [23]. SKAT has been successfully used to study variants in AD [26,27], but its associations with brain IDPs and its potentialities in the IG framework have been only partially investigated so far. In this scenario, of note is the study by Lu et al. [25] where SKAT along with group LASSO and Bayesian latent variable selection were tested on a cohort of AD subjects to identify associations between genes and nine imaging phenotypes, represented by regional volume measures. The authors demonstrated the added value of such approaches which allow accounting for the correlation among SNPs and detecting causal SNP sets, limiting the burden of multiple comparison correction. However, these promising results were achieved by analyzing each regional imaging volume separately, though, as the author recognized, these are usually correlated and their joint modeling may hold an increased power bringing additional information.

Therefore, multivariate methods represent the key to address such limitations, allowing to leverage the multiscale phenotype–genotype fingerprints while reducing the multiple testing burden, resulting in higher statistical power to identify significant associations [1,13]. Latent variable and multi-view models, for example, aim at finding a latent low dimensional space by the optimization of a target function such that the projections of the features hold some maximized joint properties. Canonical correlation analysis (CCA)-based methods have been largely applied in the IG framework in the past years, resulting into linear combinations of the two sets of variables which have maximum correlation with each other. Such approaches demonstrated high precision in assessing correlation patterns between the given features [19], for example when considering SNPs and functional MRI features [28], or in its sparse and multi-view version to establish associations between SNPs, sMRI IDPs and cognitive outcomes [29]. Partial Least Squares (PLS) analysis, which aims at maximizing at each step the covariance rather than the correlation between the latent variables, has been less frequently applied for detecting the multivariate genotype–phenotype associations. Although CCA and PLS are mathematically related, studies demonstrated that PLS may be more suitable and have improved predictive power when dealing with high-dimensional datasets, especially those with highly collinear variables common in IG experimental settings [30,31]. Interestingly, Lorenzi et al. [18] exploited PLS to uncover the genetic underpinnings of brain atrophy in AD by relying on SNPs and T1-weighted (T1-w) sMRI, demonstrating

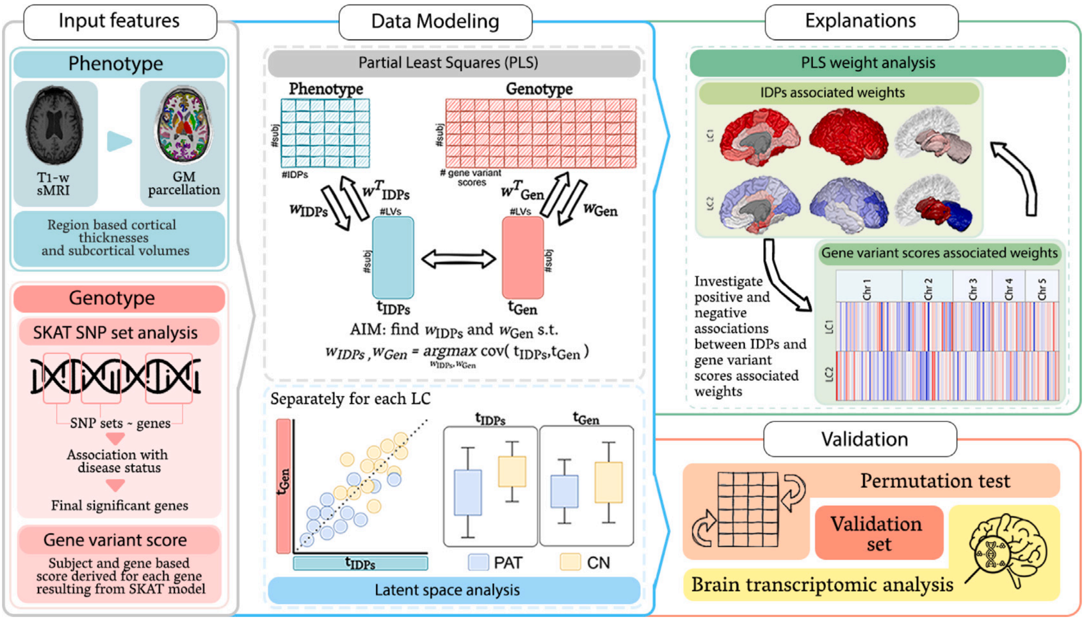
Input features
Phenotype
T1-w sMRI
GM parcellation
Region based cortical thicknesses and subcortical volumes
Genotype
SKAT SNP set analysis
SNP sets ~ genes
Association with disease status
Final significant genes
Gene variant score
Subject and gene based score derived for each gene resulting from SKAT model
Data Modeling
Partial Least Squares (PLS)
Phenotype
Genotype
#subj
#IDPs
# gene variant scores
w_{IDPs}
w^T_{IDPs}
w^T_{Gen}
w_{Gen}
#LVs
t_{IDPs}
t_{Gen}
AIM: find w_{IDPs} and w_{Gen} s.t.
$w_{IDPs}, w_{Gen} = \underset{w_{IDPs}, w_{Gen}}{argmax}\ cov(t_{IDPs}, t_{Gen})$
Separately for each LC
t_{Gen}
t_{IDPs}
PAT
CN
Latent space analysis
Explanations
PLS weight analysis
IDPs associated weights
LC1
LC2
Gene variant scores associated weights
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
Investigate positive and negative associations between IDPs and gene variant scores associated weights
Validation
Permutation test
Validation set
Brain transcriptomic analysis

Fig. 1. Overview of the proposed pipeline. Phenotype and genotype, representing region-based cortical thicknesses and subcortical volumes respectively, were given as input to Partial Least Square (PLS) modeling to model the underlying joint covariance. For each obtained Latent Component (LC) the latent space as well as the separation between patients (PAT) and controls (CN) was evaluated. Model explanations were extracted through the analysis of the PLS weights which allowed retrieving positive and negative associations between the genotype and phenotype. The model was validated through a permutation test as well as through the projection of an independent validation set on the obtained latent space which allowed to verify model generalizability. Finally, a transcriptomic analysis was performed to investigate the brain expression of the most relevant genes.

the presence of a significant link between TRIB3 and the stereotypical pattern of grey matter loss in AD. Despite these promising results, the potentialities of a classical statistical model as PLS are still under investigated. Being explainable by design, analyzing model weights allows understanding the intermediate steps and the relationship between input data and the output on top of the final outcome. In this way, such models could help disambiguating the associations between different feature sets and providing a straightforward explanation of the outcomes.

Therefore, in this work, we aimed at investigating the genetic mechanisms underlying brain atrophy in the AD continuum relying on a data-driven explainable multivariate approach (PLS) to model their joint covariation with a twofold goal: (i) exploring the association between imaging (sMRI IDPs) and genetics (SNP sets derived mutation scores) features identified in a study cohort of healthy controls and patients on the AD continuum from ADNI-3; A validation analysis was then performed with dual objectives: (i) identifying the input features driving the genotype–phenotype associations analyzing the model's weights; (ii) assessing the expression of the detected genes through a transcriptomic analysis. The resulting model was finally further validated on an independent cohort, representing the same class distribution as the discovery set.

Statement of significance:

- Problem or Issue: The association between brain phenotypes and genotype in Alzheimer's Disease is still not fully explored, in particular through multivariate analysis at a gene level and in a limited study cohort.
- What is Already Known: Conventional methods use mass-univariate techniques with complete sets of SNP to assess their association with specific brain traits. However, these methods are hindered by statistical problems and miss out on complex feature interactions, hindering the discovery of significant associations.
- What this Paper Adds: This study introduces an approach that summarizes genetic information, reducing the number of considered genetic features while incorporating gene-based data. It employs SKAT, a SNP set approach, focusing on SNPs in exon regions and creating subject- and gene-specific variant scores. The interaction between gene variant scores and comprehensive brain imaging phenotypes is then tested in a cognitively impaired cohort. Moreover, the study focuses on ADNI-3, addressing sample size limitations by proposing validation and generalization techniques, including feature distribution analysis and transcriptomic analysis of candidate genes to assess findings plausibility.

Table 1
Sociodemographic characteristics of the study cohort. Age and education are reported as years mean and standard deviation [Mean (SD)], while sex as the count of males and females individuals, respectively.

Diagnosis	Discovery set		Validation set	
	CN	PAT	CN	PAT
Count	181	62	39	15
Age, y	71.19 (6.12)	72.21 (8.88)	70.45 (6.11)	72.83 (9.56)
Education, y	17.05 (2.11)	16.11 (2.56)	16.49 (2.29)	15.67 (2.74)
Sex, m/f	73/108	38/24	14/25	11/4

2. Materials and methods

Fig. 1 shows an overview of the pipeline proposed in this work. In what follows, all the steps will be fully detailed.

2.1. Study cohort

Data used in this study were derived from the ADNI database (), in particular from the ongoing ADNI-3 phase. The ADNI was launched in 2003 as a public–private partnership led by Principal Investigator Michael W. Weiner. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Up-to-date information is available at www.adni-info.org.

Summary sociodemographic and clinical information are provided in Table 1. The considered cohort was selected based on the availability of both MRI and genetic data and ethnicity, restricting the analyses to participants with European ancestry. The final cohort comprehended

297 subjects divided into 220 healthy controls (CN) and 79 patients PAT, 19 of which were AD while the remaining were MCI subjects. The 80% of the subjects was considered as the discovery cohort, while the remaining 20% was kept for validation. A comparable proportion between PAT and CN was kept in the discovery and validation sets.

3D T1-w MRI volumes were considered for IDPs extraction (sagittal accelerated MPRAGE, TR/TE = shortest, TI = 900 ms, flip angle = 9°, Field Of View = 256 × 256 mm^2, spatial resolution = 1 × 1 × 1 mm^3, slices = 176–211). More details about the data acquisition can be found in [15]. ADNI-3 participants were genotyped using the Illumina Infinium Global Screening Array v2.

2.2. Image processing and phenotype feature extraction

The T1-w volumes were minimally preprocessed for bias-field correction (*fsl_anat* tool, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki, [32]). Subsequently, 84 anatomical regions of interest (ROIs) were extracted using FreeSurfer version 7.0 (https://surfer.nmr.mgh.harvard.edu/, [33]). The average thickness and volume were considered for cortical and subcortical ROIs, respectively. The subcortical volumes were further normalized by the estimated total intracranial volume of the respective subject. The ROIs were averaged over hemispheres resulting in 42 features to be used in the subsequent analyses. A workflow detailing the image processing steps is shown in Supplementary Fig. S1.

Moreover, as preliminary analysis, a Mann Whitney non-parametric U-test was performed to assess the group-wise differences between PAT and CN, separately for each brain feature. This allows to gain a clear insight into relations already present in the input features. False Discovery Rate (FDR) correction ($p_{fdr} < 0.05$) was applied.

2.3. Genetic processing and genotype feature extraction

Quality Control (QC) procedures were conducted on genotype data using the whole-genome association analysis toolset PLINK 1.9 [34]. SNPs and subjects were filtered out based on missingness ($geno > 0.2$, $mind > 0.1$), minor allele frequency ($MAF > 0.05$) and deviations from Hardy–Weinberg equilibrium ($hwe >$ 1e−06). QC kept 303150 SNPs out of the 759993 SNPs collected in ADNI-3. No subjects were filtered out.

GWAS analysis was performed as benchmark, including the top ten principal components from a principal component analysis (PCA) over genotype data, using age and gender as covariates.

SNP set analysis was then performed using the SKAT model [23]. Each SNP set contains the group of SNPs located in a given gene, resulting in one SNP set for each gene extracted from SKAT application and will be referred as "genes" throughout the paper. Of note, only SNPs located in the gene's exon regions were included. This led to 17 295 genes containing a total of 132 312 SNPs. SKAT was hence used to test the association between each gene and the disease status (PAT or CN) using logistic kernel-machine-based test adjusted by covariates. The R package *SKAT* was used to perform the analysis, specifying a linear weighted kernel and the same set of covariates as for the GWAS analysis.

More in details of the model, for a subject i, where $i = 1, \dots, n$, the SNP set, specific for each gene can be defined as $\mathbf{G}_i = \{g_{i1}, \dots, g_{ip}\}$, where p is the number of SNPs in the selected gene. More in detail, the state of a SNP g_{iv} is 0, if no genetic variation between the specific subject i and the reference genome is present, 1 otherwise. Given y_i the subject's disease status, the relationship between the $\mathbf{G}_i$ and y_i is given by $y_i = \alpha_0 + \boldsymbol{\alpha}'\mathbf{C}_i + \boldsymbol{\beta}'\mathbf{G}_i + \epsilon$, where α_0 is an intercept term, $\mathbf{C}_i = \{c_{i1}, \dots, c_{im}\}$ is the vector of the m covariates, $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_m\}$ is the vector of regression coefficients for covariates, $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_p\}$ is the vector of regression coefficients for the p SNPs, and ϵ is the error term (with zero mean and σ^2 variance).

Evaluating whether the gene variants influence the disease state, adjusting for covariates, corresponds to testing the null hypothesis $H_0 : \boldsymbol{\beta} = 0$, hence $\beta_{i1} = 0, \dots, \beta_{ip} = 0$. SKAT tests H_0 by assuming that for $v = 1 \dots p$ each β_{iv} follows an arbitrary distribution with mean 0 and variance $w_v \tau$, where τ is a variance component and w_v is a predefined weight for the SNP g_{iv}. The null hypothesis can be hence rephrased as $H_0 : \tau = 0$, which can be tested through a variance-component score test. This will only require fitting the null model $y_i = \alpha_0 + \boldsymbol{\alpha}'\mathbf{C}_i + \epsilon_i$. The group-wise variance-component score statistics is given by $Q = \frac{(\mathbf{y}-\hat{\boldsymbol{\mu}})' K (\mathbf{y}-\hat{\boldsymbol{\mu}})}{2}$, where K is the weighting linear kernel, $\hat{\boldsymbol{\mu}} = \hat{\alpha_0} + \mathbf{C}\hat{\boldsymbol{\alpha}}$ is the predicted mean of $\mathbf{y}$ under H_0 and $\hat{\alpha_0}$ and $\hat{\boldsymbol{\alpha}}$ are estimated under the null model by regressing $\mathbf{y}$ on only the covariates $\mathbf{C}$. More in detail, K, is a $n \times n$ matrix where each entry measures the genetic similarity between two subjects i and i' in the gene given the p SNPs.

To derive a p-value for the considered gene, SKAT tests if Q follows a mixture of χ^2 distributions. In our analysis a gene is considered as significant if its associated p-value is $\leqslant 0.05$. Of note, Supplementary Algorithm 1 reports the SKAT algorithm listing.

Once the significant genes were obtained through SKAT, a function to map the population significant genes to a subject specific measure was proposed. In detail, for a subject i and for a significant gene G resulting from SKAT , a gene-based variant score $\eta_i(G)$ was extracted representing the total mutation score in G.

No distinction for diploid variations at the same locus was considered. The gene variant score of G for each subject i and each significant gene resulting from SKAT, was defined as

$$\eta_i(G) = \frac{\sum_{v=1}^{p} g_{iv}}{p}.$$

A workflow detailing the genetic processing steps is shown in Supplementary Fig. S2. Of note, as it has been done and described for the phenotype, a Mann Whitney non-parametric U-test was performed on the gene variant scores to assess the group-wise differences between PAT and CN, followed by FDR correction ($p_{fdr} < 0.05$).

In order to better analyze the genes resulting from SKAT analysis, their association with the disease state was furtherly assessed using Hetionet [35] and REACTOME pathway analysis (R package ReactomePA, [36]). In detail, Hetionet is an open-source biomedical graph database that combines the information from 29 public databases into a single resource. It contains 47 031 nodes of 11 types (e.g. genes, diseases, pathways, compounds) and 2 250 197 edges of 24 types (e.g. upregulates/downregulates, interacts). All the significant SKAT genes were searched inside Hetionet in order to retrieve their eventual link with AD. REACTOME was additionally used to perform enrichment analysis starting from the full set of significant SKAT genes. Significant pathways were selected based on the associated FDR-adjusted p-value ($p_{fdr} < 0.2$). Among these, pathways associated with AD in Hetionet were selected.

2.4. PLS analysis

Phenotypes and genotypes were organized in two separate data matrices, **X** and **Y**, respectively, subsequently divided in discovery and validation sets. To ensure that the differences in thickness, volume magnitudes or SKAT scores would not dominate the statistical model, the **X** and **Y** data matrices were z-scored column-wise, by subtracting the mean from each column and dividing by the standard deviation of that column. Moreover, the association between the most commonly considered covariates such as age, sex, APOE and years of education was tested through Spearmann correlation with both the genotype and the phenotype. Results showed that only age was correlated with the phenotype, while no association was recorded for the other variables. The influence of age was hence regressed out from the phenotypes only.

With the goal of modeling the joint variation between morphometric IDPs and gene variant scores in our discovery cohort, the PLS model was applied following [18,37,38]. Among the numerous versions of PLS, the symmetric PLS formulation computed using the NIPALS (Nonlinear Iterative Partial Least Squares) algorithm [39] was considered.

Within this setting, PLS is intended to estimate the latent components (LCs) that maximize the global covariance between the two input modalities. More in detail, this model aims at identifying the vectors $\mathbf{w}_{x\_opt}$ and $\mathbf{w}_{y\_opt}$ such that the covariance between the projections of the two input variables, $\mathbf{X}\mathbf{w}_{x\_opt}$ and $\mathbf{Y}\mathbf{w}_{y\_opt}$, is iteratively maximized:

$$\mathbf{w}_{x\_opt}, \mathbf{w}_{y\_opt} = \underset{w_x, w_y}{\mathrm{argmax}}\left(\frac{\mathbf{w}^T{}_x \mathbf{S}_{\mathbf{XY}} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{w}_x}\sqrt{\mathbf{w}_y^T \mathbf{w}_y}}\right) \tag{1}$$

where $\mathbf{S}_{\mathbf{XY}}$ is the cross-covariance matrix between the feature matrices $\mathbf{X}$ and $\mathbf{Y}$.

The optimal number of PLS LCs was chosen by calculating the data variability explained by each of them based on model singular values. The ratio of the singular value to the sum of singular values from the decomposition was used to threshold the number of components in order to retain the 60% of explained data variability.

LC related projection scores t_x and t_y were derived separately for genotype and phenotype by multiplying the respective input by the LC associated weights ($\mathbf{t}_x = \mathbf{X}\mathbf{w}_{x\_opt}$, $\mathbf{t}_y = \mathbf{Y}\mathbf{w}_{y\_opt}$). The group-wise Mann Whitney non-parametric U-test was then applied on the projection scores to assess group-wise differences between PAT and CN. Only the associations generated by LCs showing significant separation between the two groups on both the phenotype and genotype were retained for further analysis. A workflow detailing the PLS modeling is shown in Supplementary Fig. S3.

2.5. PLS explainability

PLS model belongs to the class of the 'white box' models, hence models for which explanations are immediately available. Following Eq. (1), in each LC, each input feature is given a weight according to its relative importance for describing the global multimodal relationships across the input features. The magnitude of the associated weight directly reflects the importance of each feature in the common latent space, while its sign indicates the direction of the latent association, direct or inverse, also referred to as correlation or anticorrelation. This sign does not necessarily entail an effective increase or decrease of a particular feature's input value in a group of subjects compared to the other. Rather, it simply describes the association between features as found by the PLS model. The preliminary analysis on the input, described in Sections 2.2 and 2.3, will aid the interpretation of the obtained associations.

2.6. PLS validation

A permutation test based on the obtained singular values was performed to assess the significance of the model defined on the discovery set [40]. In brief, the test checked whether the singular values associated to each LC were higher than the ones obtained by randomly permuting all rows of the phenotype matrix (1000 permutations were used).

The generalization capability of the PLS model was then tested on the unseen validation group by statistically assessing the ability of the estimated PLS components in splitting patients and controls through group-wise comparison of the projection scores in the latent space (Mann Whitney non-parametric U-test). PLS analysis and validation were performed using Python, relying in particular on the `scikit-learn` library [41]. The code is publicly available at https://github.com/fcrucian/PLS_ImagingGenetics/.

2.6.1. Transcriptomic analysis

Finally, a transcriptomic analysis was performed based on the Human Protein Atlas (HPA) database (,[42]). The HPA provides normalized transcript per million (nTPM) expression values within 13 brain regions based on RNAseq analysis of 1324 samples from several donors. Each of the most relevant genes either belonging to Hetionet database or resulting both from the [enrichment] analysis and the PLS model LCs analysis was checked for expression in brain tissues. Of note, the HPA does not have a reference template for brain regions definition. Their transcriptomic measures were hence aggregated in order to match the Desikan-Killiany Atlas considered in our study for the IDPs definition. A full matching was not however possible due to missing information in HPA database, resulting in 39 regions out of 42 having transcriptomic data. Of note, the expression values were normalized in the range [0, 10].

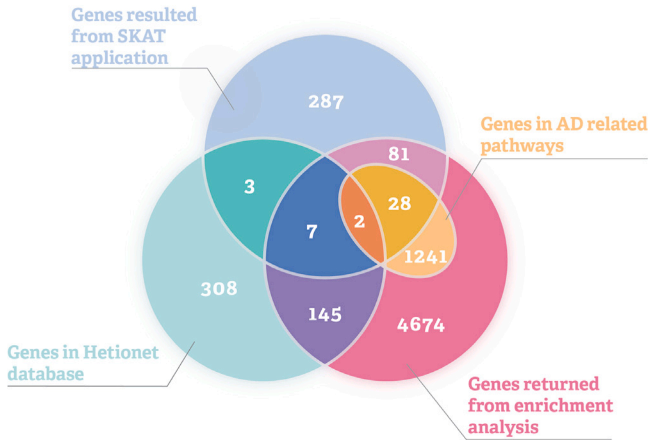
Genes resulted from SKAT application
Genes in AD related pathways
Genes in Hetionet database
Genes returned from enrichment analysis
287
81
3
28
7
2
1241
308
145
4674

Fig. 2. Overview of the gene's grouping. The Venn diagram is based on four main sets representing genes resulting from Sequence Kernel Association Test (SKAT) analysis (light blue), genes in Hetionet database (turquoise-green), genes belonging to the significant pathways returned from the enrichment analysis (pink) and the subset of the latter representing the genes belonging to Alzheimer's Disease (AD) related pathways (yellow). Darker colors are used to represent the intersections between such principal clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results

In what follows, the main results obtained in this study are presented. GWAS and SKAT analyses on the genetic data will be followed by the results of preliminary statistical analyses on the genotype/phenotype variables. Then, the PLS modeling findings will be presented starting from the analysis of the latent projection scores, moving to the genotype–phenotype interactions and ending with the transcriptomic analysis outcomes.

3.1. GWAS and SKAT results

GWAS and SKAT analyses were conducted on the complete study cohort after a QC preprocessing in order to identify genotype associations from case-control data (PAT and CN subjects) as described in Section 2.3. GWAS failed to discover significant associations between individual SNPs and disease status, possibly due to the limited sample size and low prevalence of disease [43] in the considered cohort. The related Manhattan plot is shown in Supplementary Fig. S4.

The SKAT SNP set analysis revealed 408 significant genes (p-value $\leqslant$ 0.05). These were almost equally distributed in all chromosomes, though a higher predominance could be noted in chromosome 6 (48 genes, 12% of significant genes) and chromosome 1 (33 genes, 8% of significant genes). A schematic overview of the different gene subsets retrieved in our study can be found in Fig. 2 , and will be detailed in what follows. 12 SKAT significant genes were found associated with AD in Hetionet, and we refer to these as "Hetionet genes". In details: PTGS2 and DPYD (Chr1), TF (chr3), PPARGC1 A (Chr4), CDH12 (Chr5), VEGFA (Chr6), LPL (Chr8), CHAT and ABCC2 (Chr10), BDNF (Chr11), AKAP13 (Chr15), CYP2D6 (Chr22).

Table 2

Significant pathway associated with Alzheimer's Disease (AD) in Hetionet. REACTOME was used to conduct enrichment analysis. For each pathway is reported the Reactome ID, the pathway name, Sequence Kernel Association Test (SKAT) genes included in the pathway (Hetionet genes are highlighted in bold), *p*-value and false discovery rate adjusted *p*-value.

ID	Pathway	Genes	*p-value*	*p-fdr*
R-HSA-211999	CYP2E1 reactions	**CYP2D6**/CYP2E1/CYP2C9	0.001	0.08
R-HSA-211897	Cytochrome P450	**CYP2D6**/CYP2E1/CYP2C9/CYP4V2/CYP11B1/CYP4F12	0.005	0.12
R-HSA-211859	Biological oxidations	FMO2/**CYP2D6**/CYP2E1/GSTM5/CYP2C9/EPHX1/ CYP4V2/MAT1A/UGT2B4/CYP11B1/CYP4F12/MTARC1	0.007	0.14
R-HSA-112316	Neuronal System	**CHAT**/RPS6KA2/KCNA7/SLC1A2/GABRG2/KCNH5/ CACNA1A/KCNMB1/CASK/SLC6A1/KCNJ6/KCNAB1 CHRNA5/KCNMB3/NRXN1/ABAT/GRIN3A/GABBR1	0.010	0.17

Enrichment analysis identified 53 significant pathways (FDR adjusted *p*-value < 0.2), as reported in Supplementary Figure S6. Among these, four pathways were associated with AD in Hetionet, reported in Table 2 together with Reactome ID, SKAT genes included in each pathway, *p*-values and adjusted *p*-values. The Hetionet genes included in these pathways (CUP2D6 and CHAT) are highlighted in bold. While the association between the *Neuronal System* pathway and AD is clear, for the other three pathways the relationship is highlighted hereafter. *Biological oxidation* has been demonstrated to be associated with cell toxicity in various neurodegenerative disorders such as AD or Parkinson's Disease. An accumulation of nucleic acid oxidation indicates a decreased capacity to repair the nucleic acid damage [44]. Furthermore, *CYP2E1* reactions pathway is closely associated with *Biological oxidation*. CYP2E1 gene is involved in oxidative stress and can cause cell death [45]. Finally, *Cytochromes P450* in the corresponding pathway constitute a superfamily of enzymes that catalyze the metabolism of drugs. Polymorphisms in cytochrome P450 genes may affect the enzyme catalytic activity and have been associated with AD in several studies [46,47].

3.2. Phenotype and genotype preliminary analysis

The preliminary analysis on the phenotype, aiming at assessing whether any between-group significant difference was present in the original space, revealed selective alterations surviving the FDR correction. Among the cortical IDPs, temporal regions were the most significant (p-value $\leqslant$ 1e−05 for enthorinal cortex, middle temporal gyrus and temporal pole; p-value $\leqslant$ 1e−04 for fusiform gyrus, inferior temporal cortex, parahippocampal gyrus, superior temporal gyrus) followed by few parietal regions (precuneus and insula, p-value $\leqslant$ 1e−04) and by banks of the superior temporal sulcus and inferior/superior parietal gyri (p-value $\leqslant$ 1e−03). Moving to the subcortical IDPs, amygdala was the most significant one (p-value = 3.56e−08), followed by hippocampus and accumbens recording *p*-values of 2.15e−07 and 1e−04, respectively. All the statistics revealed a decrease of the measured features in PAT compared to CN. No significant differences were found for the remaining phenotype features. Moving to the genotype, 60 gene variant scores revealed significant differences between PAT and CN. ChrX had the highest percentage of significantly different genes (42%). It was followed by Chr4, Chr22, Chr1, Chr17 and Chr2 which showed a percentage of significant genes above the 15%. However, no comparison survived the FDR correction.

3.3. PLS analysis

The X and Y matrices for PLS computation had dimension number of subjects [243 for the discovery and 54 for the validation] × number of respective features [42 IDPs for X and 408 gene variant scores for Y]. The PLS model on the discovery set returned a total of 14 LCs needed to explain at least the 60% of data variability, the first accounting for the 12%, and the others monotonically decreasing till the 3%. In what follows, out of the 14 LCs, we will focus on those components featuring significantly different project scores across groups. Of note, the permutation test confirmed the significance of the model resulting in a *p*-value = 0.001.

3.3.1. Latent space and projection scores

Among the 14 LCs, the 1st (LC1), the 2nd (LC2) and the 5th (LC5) where the ones showing significant differences between the projection scores of PAT and CN groups for both the genotype and phenotype in the discovery set. Such LCs accounted for the 12%, 7% and 4% of data variability, respectively. Fig. 3 shows the latent space spanned by such LCs, as well as the distribution of the related projection scores, separately for imaging and genetics and for both the discovery and the validation sets. Focusing on the discovery set, high correlation was present between genotype and phenotype projections for all the considered LC (Pearson correlation coefficient equals to 0.82, 0.80 and 0.77 for LC1, LC2 and LC5, respectively), while a clearer separation between classes was present in LC2 and LC5, compared to LC1. Moving to the differences in projection scores between PAT and CN, LC1 showed a p-value $\leqslant$ 1e−02, namely *p*-value of 0.002 and 0.001 for phenotype and genotype projection scores, respectively. LC2 and LC5, despite accounting for a minor data variability, appeared to be more powerful in discriminating PAT and CN. LC2 showed strong significant differences for both phenotype and genotype, with a *p*-value of 1e−04 and 4e−05, respectively. A similar trend was observed for LC5 which showed *p*-values of 2e−04 (phenotype) and 2e−08 (genotype).

The projection of validation data on the generated latent space showed a similar distribution of PAT and CN patterns as the discovery set and confirmed some of the significant differences recorded for the discovery set. In detail, for LC1, significance (*p*-value = 0.014) was found also on the validation set for the phenotype projection. The validation set on LC2 showed a significant difference for the genotype (*p*-value = 0.008), also found in LC5 (*p*-value = 0.010). For both LC2 and LC5 the phenotype in the validation set showed a trend toward the significance (*p*-values of 0.072 and 0.061, respectively), though still not reaching it.

3.3.2. Genotype–phenotype relevance and associations

Fig. 4 shows the phenotype (imaging) weights from the PLS model, separately for LC1, LC2 and LC5. Overall, distinct patterns emerge among the considered LCs, revealing the unique brain structure modulations explained by each of them in the considered study cohort. Specifically, cortical regions received high positive weights in LC1, suggesting their predominant role in shaping this component, with a focus on frontal (parso percularis, rostral middle and superior frontal gyri, precentral gyrus), temporal (inferior, middle and superior temporal gyri, fusiform gyrus, bankssts) and parietal areas (supramarginal gyrus, inferior parietal gyrus). Conversely, subcortical regions received low importance. Conversely, LC2 assigned high (positive) weights to the subcortical features, showing an anticorrelation between them and the cortical ones. The most important volumes were hippocampus, amygdala, basal ganglia (putamen, globus pallidus, caudate, and accumbens), and thalamus. These resulted to be positively correlated with the entorhinal cortex and temporal pole, similarly featuring high positive weights, and anticorrelated particularly with the cerebellum and rostral middle frontal gyrus. Finally, LC5 highlighted a strong separation between frontal regions (frontal pole, parsorbitalis, parstriangularis, cingulate gyri, in particular caudal anterior, rostral anterior

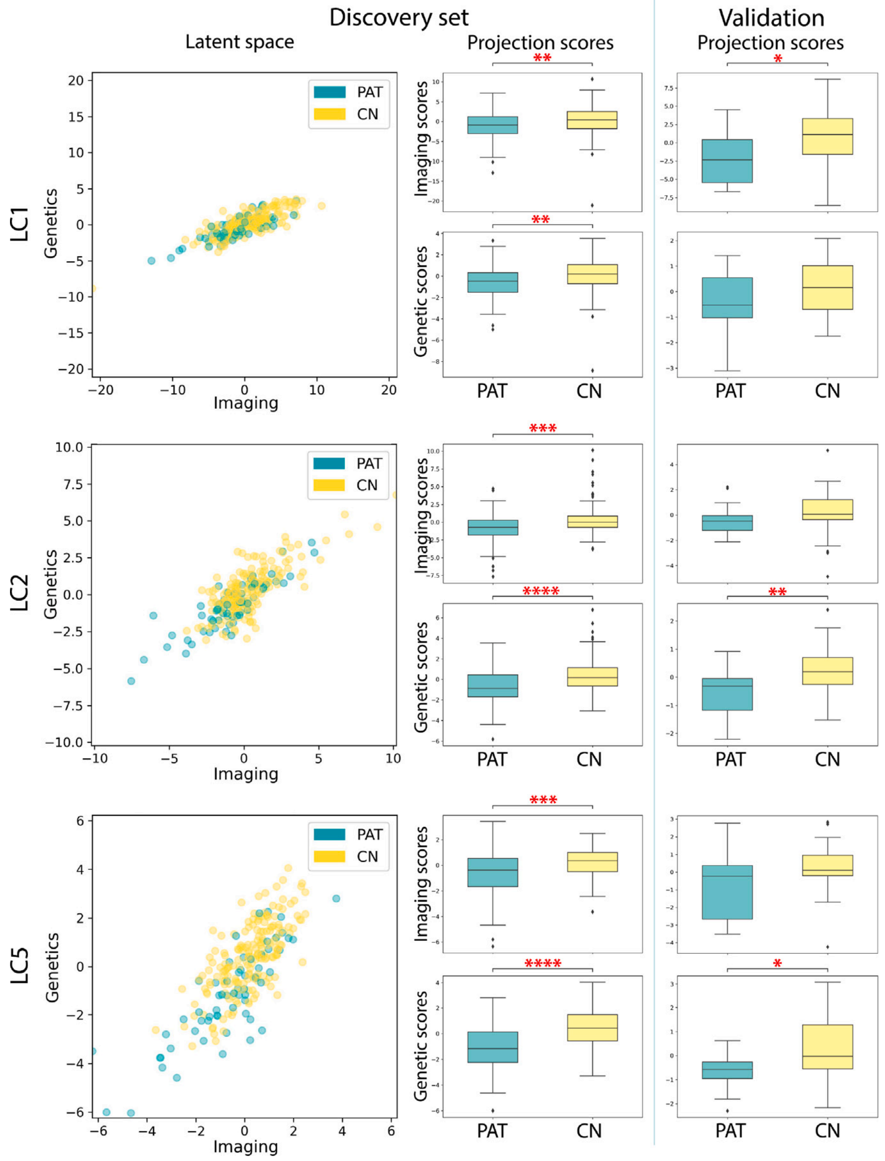

Fig. 3. Latent space and projection scores boxplots. The latent space projections on the Partial Least Squares (PLS) latent components (LCs) showing a significant difference between patients (PAT, blue) and controls (CN, yellow) are shown in rows. The projection scores for phenotype and genotype separately are then reported for both the discovery set and the validation set (columns). Significant differences between CN and PAT projections, as derived from Mann Whitney non-parametric U-test, are highlighted with red asterisks (*, **, ***, **** refers to *p*-values $\leqslant 0.05, 1e-02, 1e-03, 1e-04$, respectively). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and posterior), medial temporal lobe (entorhinal cortex, parahippocampal gyrus), subcortical volumes (negative weights) with temporal lobe regions (temporal pole, middle temporal gyrus).

Moving to the genotype (gene variant scores), Supplementary Fig. S7 shows the associated PLS weights, separately for LC1, LC2 and LC5. Despite the relevance pattern resulted quite uniform across chromosomes, some differences emerged. More in detail, in LC1 Chr2, Chr3, Chr11, Chr21 and ChrX showed the genes featuring, on average, the highest negative weights in anticorrelation with Chr18 which had instead the highest positive ones. In LC2, the chromosomes featuring the highest positive weights were Chr7, Chr18, opposed to Chr17, Chr19, ChrX. Finally, for LC3, Chr12 showed the highest positive and negative weights. Chr4 and Chr11 (positive weights) were in negative correlation with several chromosomes (Chr9, Chr18, Chr20, Chr21, Chr22, ChrX).

In order to better emphasize the association between phenotype and genotype, Fig. 5 reports an heatmap illustrating the relative PLS weights for each feature and component. For ease and clarity, imaging features were grouped by cortical (Cort) and subcortical (Subc) regions, while genes were grouped by their position on chromosomes. An empirical threshold to retain only the weights higher than the 75th percentile of the respective distribution was applied. Moreover, Table 3 highlights the most relevant associations commented below.

A first macro-analysis was performed on the genetic side, by analyzing the global importance of each chromosome. In particular, the percentage of SKAT genes above the threshold normalized by the total number of SKAT genes in a given chromosome was computed. Results showed that the chromosomes featuring the highest percentage of relevant genes were Chr18, Chr21 and ChrX for LC1, including the 40% (Chr18 and Chr21) and 42.8% (ChrX) of relevant SKAT genes. Chr7, Chr17, Chr18 and Chr19 had the 44.4%, 45%, 40%, 42.8% and 40% respectively covered by the most relevant SKAT genes in the LC2. Finally, for the LC5, the Chr4, Chr9, Chr12 and ChrX resulted in percentages of 41.2, 42.8, 40.9, 42.8 of SKAT genes with the associated weights above the threshold, respectively.

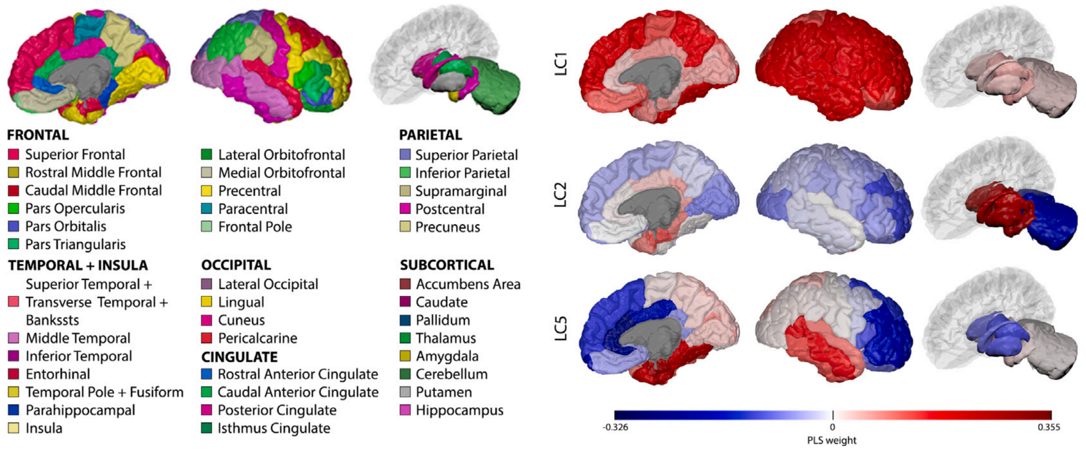

Fig. 4. Significant Partial Least Squares (PLS) components' weights for the phenotype. The three latent components (LCs) are shown in rows (right). Positive weights are shown in red, while negative ones are in blue. Reference Desikan–Killany atlas [48] highlighting the regions considered in this study is shown on the left. Drawings generated using BrainPainter [49]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

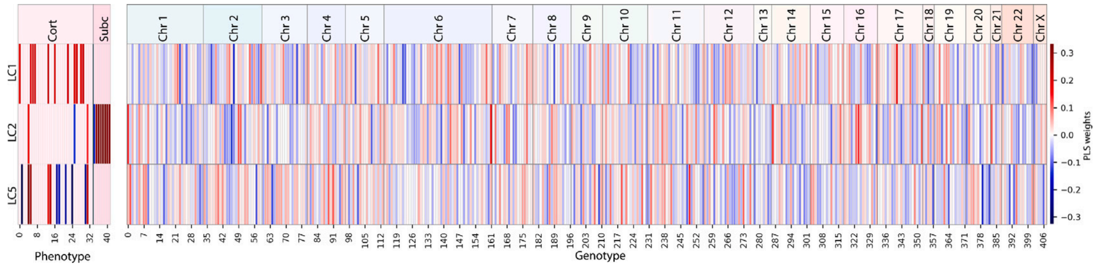

Fig. 5. Association between phenotype and genotype. Heatmap of the Partial Least Squares (PLS) weights for the selected latent components (LC, rows), thresholded over the 75th percentile of the respective distribution. Background shades highlight cortical (Cort) and subcortical (Subc) features for phenotype, and different chromosomes (e.g. Chr1) for genotype. The corresponding feature name lists can be found in Supplementary Fig. S5. Positive and negative PLS weights are shown in red and blue, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Most relevant association found though the Partial Least Squares (PLS) model in each significant latent component (LC). The features are ordered in terms of associated weight, positive and negative weights are highlighted in red and blue, respectively.

LC1		LC2		LC5	
Imaging	Genetics	Imaging	Genetics	Imaging	Genetics
supramarginal gyrus	MFSD6L	hippocampus	PHF14	entorhinal cortex	BCAS1
middle temporal gyrus	RIF1	putamen	RFWD3	caudal anterior cingulate	GLT6D1
inferior parietal gyrus	ATP6V1G2	globus pallidus	MORN1	rostral anterior cingulate	TMPRSS15
superior frontal gyrus	NFASC	caudate	TEP1	fusiform gyrus	COL6A3
inferior temporal gyrus	FBDXD43	accumbens	HNMT	frontal pole	CEP164
precentral gyrus	KCNA7	amygdala	BDNF	parahippocampal gyrus	TF
rostral middle frontal gyrus	KCNJ6	thalamus	CACNA1A	pars orbitalis	CHAT
pars opercularis	CYP11B1	cerebellum	FMO2	temporal pole	MAT1A
bankssts		entorhinal cortex	EPHX1	posterior cingulate	CYP2C9
fusiform gyrus		rostral middle frontal gyrus	CYP2D6	pars triangularis	SLC1A2
superior temporal gyrus		temporal pole		middle temporal gyrus	KCNH5
					CYP4F12

More in depth of the relevant genes in each component, the top 5 genes showing the highest importance for LC1 were RIF1, ATP6V1G2, NFASC and FBXO403 (negative weights, Chr2, Chr6, Chr1, Chr8 respectively), in anticorrelation with MFSD6L (Chr17). The first four genes were also found to be anticorrelated with the most relevant cortical features on the phenotype, namely the frontal and temporal regions described in the previous paragraph. No weights higher than the 75th percentile threshold were recorded for the Hetionet genes in LC1. However, among the SKAT genes belonging to the four AD pathways, KCNA7 and KCNJ6 (negative weights, Chr19 and Chr21) were correlated with RIF1, ATP6V1G2, NFASC and FBXO43, inheriting the related relation with the phenotypic counterpart detailed above. On the other end, CYP11B1 (pathway R-HSA-211859, Chr8) was correlated with MFSD6L, hence in anticorrelation with the relevant LC1 cortical thickness features.

Moving to LC2, the top 5 most important genes were HNMT (negative weight, Chr2), in anticorrelation with PHF14, RFWD3, MORN1 and TEP1 (Chr7, Chr16, Chr1, Chr14) on the genetic side as well as anticorrelated with the subcortical volumes (except for the cerebellum) for the imaging features. Moreover, PHF14, RFWD3, MORN1 and TEP1 in opposition with HNMT, showed a correlation with the thickness of cerebellum cortex and rostral middle frontal gyrus among the others.

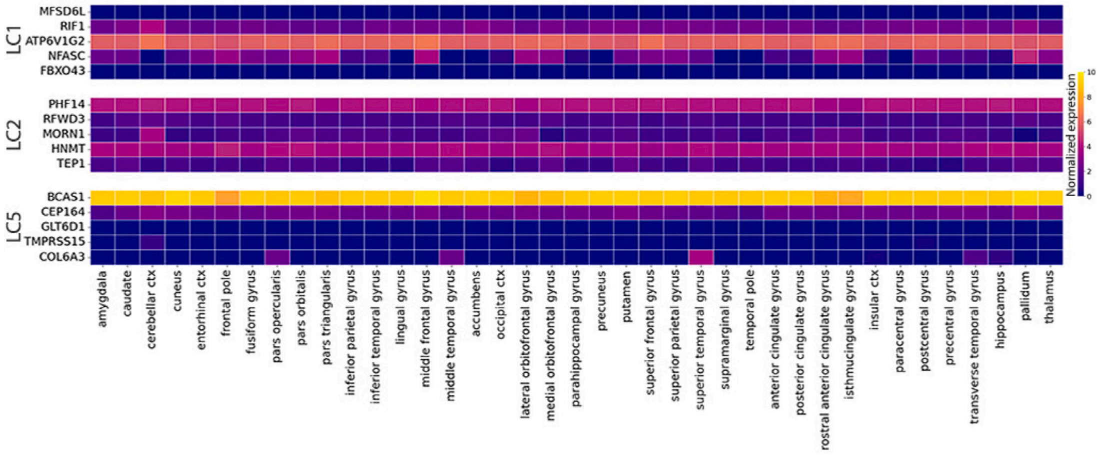

Fig. 6. Gene expression profiles for the top five genes in each significant Partial Least Squares (PLS) latent component (LC). Expression values are normalized in the range [0, 10] and grouped in the same regions as T1-weighted parcellation for the available regions.

The Hetionet genes relevant in LC2 were the BDNF and CYP2D6 (positive weights, Chr11 and Chr22). These were further correlated with the relevant subcortical regions (except for the cerebellum) in LC2. Analyzing the genes belonging to the AD pathways, CACNA1A, FMO2, EPHX1 and CYP2D6 (negative weights, pathways R-HSA-112316 and R-HSA-211859, Chr19, Chr1, Chr1, Chr22) had associated weights over the threshold in LC2. More in details, CACNA1A, FMO2 and EPHX1 were correlated with HNMT and hence inheriting its relation with the phenotype. CYP2D6 was instead found in correlation with such aforementioned imaging features, as well with the genes in the top positions in terms of weights (PHF14, RFWD3, MORN1, and TEP1) and BDNF.

Finally, the top 5 genes for the LC5 were BCAS1, GLT6D1, TMPRSS15, COL6A3 (negative weights, Chr20, Chr9, Chr21, Chr2) and were anticorrelated with CEP164 (Chr11). The latter was correlated with entorhinal cortex, fusiform gyrus and temporal pole, while the others were in correlation mainly with cingulate gyri and frontal pole. Among the Hetionet genes found in the SKAT set, CHAT, TF and BDNF (positive weights, Chr10, Chr3, Chr11) had an associated weight higher than the $75^{t}h$ percentile threshold and were found in correlation between each other in this LC, as well as with entorhinal, fusiform gyrus and temporal pole in the phenotype. An anticorrelation was found instead with cingulate gyri and frontal pole. Of note, LC5 was the component in which most of genes belonging to AD-related pathways were found with the highest weights, namely CHAT, MAT1 A, CYP2C9, SLC1A2, KCNH5 and CYP4F12 (positive weights, pathways R-HSA-211999, R-HSA-211897, R-HSA-211859, R-HSA-112316, Chr10 for the first three, Chr11, Chr14, Chr19). All of them, with the exception of KCNH5, were correlated with TF, BDNF and CEP164, inheriting their association with phenotype, as well as with cingulate gyri and frontal pole.

3.3.3. Transcriptomic analysis

The transcriptomic analysis revealed that part of the relevant genes discussed in the previous sections was also expressed in brain tissues. Fig. 6, Supplementary Figures S8 and S9 represent the normalized expression profiles for the top 5 genes, Hetionet genes and genes belonging to AD pathways in each significant component, respectively. More in detail, among the top 5 genes in LC1, RIF1, ATP6V1G2 and NFASC were generally expressed in brain. The NFASC and RIF1 showed also a peculiar expression pattern, the former featuring markedly higher levels in frontal lobe, cingulus and globus pallidus while the latter showing higher expression in the hippocampus. All the top 5 genes of LC2 were expressed in brain, with PHF14 and HNMT showing the highest values overall the brain tissues. Of note, the MORN1 was found to be highly expressed in cerebellar cortex compared to the other regions. Finally, BCAS1 and CEP164, prominent in LC5, showed a brain-wide expression, while COL6A3 was found highly expressed especially in superior and middle temporal gyri. Among the genes involved in the AD pathways and showing an associated PLS weight above the threshold of the $75^{t}h$ percentile, high expression levels were shown by KCNJ6, relevant in LC1, the EPHX1 and CACNA1A (LC2), and the SLC1A2, CYP4V2, KCNH5 (LC5). The latter in particular was expressed in entorhinal gyrus as well as in thalamus and middle frontal gyrus. Finally, all the Hetionet genes found above the threshold were found to be expressed in brain with the exception of CYP2D6. A particularly high expression was found for TF (LC5).

4. Discussion

In this work, we addressed the twofold objective of modeling the associations between brain imaging and genetics in patients on the AD continuum, highlighting the genes and the brain regions leading them. This was achieved by modeling the joint covariation between 42 region-based morphometric measures, detecting possible patterns of cerebral atrophy, and 408 gene variant scores, calculated for the significant genes derived from the SKAT SNP set approach. A well assessed model, the PLS, was applied to this aim, allowing to obtain explainable latent phenotype–genotype associations. Compared with previous approaches, this study firstly proposes a method to summarize the genetic information, which allowed to exploit gene-based information while reducing the number of genetic features considered, overcoming the limitations inherent to GWAS or single SNP analyses. This was achieved by exploiting SKAT as SNP set approach (considering only SNPs located in the exon regions) and then projecting back the results to a subject level by computing a subject and gene specific variant score representing how varied the gene was for the specific subject compared to a reference genome. In this way, gene variant score allowed the characterization of each significant gene highlighted by SKAT with a single value, which could be then used along with imaging variables in multivariate IG models. To the best of our knowledge, the interaction between gene variant scores and a complete set of brain structural imaging phenotypes has not been yet investigated in a cognitive impaired cohort, though can convey novel and more meaningful information compared to considering each single SNP or summary risk scores at a time. Moreover, we focused on ADNI-3 to investigate the potentialities of this dataset, as this is still under-investigated when considering IG associations mainly due to sample size limitations inherent to the available genetic data for this study cohort. Indeed we considered 297 individuals, divided into healthy CN and PAT (either MCI or AD) and further split into discovery and validation cohorts (80% and 20%, respectively). We finally proposed validation and generalizability techniques suitable for a small study cohort, among which a preliminary statistical analysis on the input feature distribution to better interpret and validate the obtained results and the transcriptomic analysis on the obtained candidate genes to assess the plausibility of our findings.

Summary of main findings. In terms of imaging and genetic variables, while the phenotype features are here represented by well-known morphometric measures for regions that have been proven to be involved at different levels in the neurodegeneration process typical of the AD continuum [6,7], the significant genes resulting from SKAT method revealed twelve genes belonging to the Hetionet database. Hence, this method, though applied in a somehow limited cohort where conventional GWAS failed, was able to retrieve well-known genes known for their association with AD. Moreover, four out of the significant pathways obtained through the enrichment analysis on the significant genes were as well associated with AD in Hetionet. The joint multivariate modeling between imaging and genetics relied on the PLS, an explainable model, which allowed to derive significant genotype–phenotype associations, as verified by permutation testing, and returning LCs in which a clear and significant difference between the PAT and CN projections scores could be recorded. In the LCs encompassing significant differences between PAT and CN, the relevant genotype–phenotype associations can be summarized as follows: (i) The correlation between the EPHX1 variant score (*Biological Oxidation* pathway), whose role in neurodegeneration is highly investigated and strongly supported by previous findings, and a decrease in subcortical volumes, typical of neurodegeneration. This result was also confirmed by the expression analysis which highlighted the EPHX1 to be widely expressed in brain; (ii) The correlation between the BCAS1 variant score and a significant decrease in temporal lobe thickness (PAT < CN). This gene is indeed involved in the process of myelination, particularly investigated in the dentate gyrus, part of the temporal lobe; (iii) Multiple associations, for which further exploration is needed, between the decrease in cortical thickness or volume of well known brain regions involved in AD continuum with genes whose function is still unclear, though preliminary related to neurodegeneration and which will be further detailed below.

IDPs and genes preliminary analyses. Our investigation started from a preliminary analysis of the input features, where we tested whether significant between-group differences were present, considering each feature separately from the others and relying on the Mann Whitney non-parametric U-test. On the phenotype, the test highlighted differences in thickness or volumes for well established brain regions affected by AD. Indeed, while regions along the hippocampal pathway were found to be affected by atrophy in the early stages of the disease, temporal, parietal and frontal neocortices emerge at later stages [6,8]. Moreover, a very recent systematic review on prospective biomarkers of AD [50] performed meta-analyses based on random-effect models on 84 articles, concluding that 20 biomarkers were globally associated with AD progression. Among them, hippocampal, entorhinal cortex and middle temporal lobe volumes resulted as promising prospective sMRI biomarkers for AD progression. All the aforementioned regions resulted as significantly different between PAT and CN also in our cohort, with reduced volumes in patients as expected and in line with the neurodegeneration atrophy pattern. Moreover, interestingly, such regions were also among the most relevant in the association with genetics computed through our PLS model. On the genetic side, no genes survived the FDR correction, probably due to the high number of comparisons to be considered (408). However, when no corrections were applied, 60 genes resulted significantly different between PAT and CN. Four of them, namely RIF1, PHF14, KCNH5 and HNMT were then found among the most relevant ones in the PLS LCs, hence holding an important role in the latent space definition. Of note, differences in both directions (PAT < CN and PAT > CN) were found for such gene variant scores, suggesting that variants in some genes could lead to increased resistance to the disease. However, it is important to note that they could eventually represent biases in considering all SNPs in the genes.

PLS models significance, validation and generalizability. Aiming at analyzing the multivariate association between the complete set of genetic and imaging features typical of IG studies, latent view methods such as CCA or PLS have gained increased popularity. An extensive review of the models is available in the literature [1,13]. Focusing on PLS, which is a tried-and-true technique for multivariate analysis, this method has been used with promising results to establish a connection between brain atrophy and individual SNPs from AD patients in a recent paper by Lorenzi and colleagues [18], revealing a strong link between the TRIB3 gene and the characteristic pattern of grey matter loss in such disease. They used the entire set of SNPs for the genotype and sMRI characteristics as IDPs, demonstrating the generalizability of their model in a separate cohort. The same strategy was used by Casamitjiana and colleagues [51], who were able to stratify the early stages of AD in the PLS latent space by utilizing T1-w features and the amounts of the biomarkers t-tau, p-tau, and amyloid-beta in the cerebrospinal fluid. Finally, in a previous preliminary work this approach allowed us to uncover significant associations between brain atrophy and 14 AD-related PRSs, possibly revealing different associations for different AD subtypes [38]. Interestingly, thanks to its advantages in scalability and its ability to face collinearity, PLS is starting to be applied also in the imaging transcriptomics field [52], opening new opportunities to investigate how the spatial patterns of gene expression relate to anatomical variations in brain structure and function in both health and disease. When applying the PLS model, a solution that maximizes the covariance between latent space projection is always obtained, making the validation of the results stringent. To address the issue of limited sample size in the cohort, the analysis started by examining the singular values that define the LCs through the implementation of a permutation test. The rows of the X matrix, representing the phenotype, were randomly permuted in order to break any existing connection between the IDPs and the gene variant scores. The singular values obtained from the permuted inputs were then compared with the true ones and it was hence possible to verify that a significant difference between the random singular values distribution and the true ones was present. This confirmed the relevance of the genotype–phenotype association described by the LCs associated with such singular values. Secondly, the observation of the latent space, where, separately for each LCs, the IDPs projection was plotted against the gene variant score one, allowed to assess whether the solution provided by the PLS effectively found a covariance between features. Finally, the generalization of the model was investigated through the projection on the obtained latent space of set of subjects set aside from the full cohort.

Role of the input feature definition. Besides the study of the significance and generalizability, one main limitation of these approaches is that, in order to handle a large number of input features, which is always the case when considering the full set of SNPs or the total number of voxels for imaging, a large number of observations is needed, hence they are poorly applicable when dealing with small cohorts. To overcome this limitation, on the imaging side, features computed over brain regions, rather than single voxels, were applied to check the association with genetics. These metrics on sMRI data, could represent grey matter volumes for a set of regions of interests [53], or local grey matter density extracted through voxel-based morphometry and then grouped by target regions [54]. A complete overview on the commonly used IDPs in IG studies can be found in [1]. Region-based volumes and thicknesses were indeed considered in the present study. On the genetic side, features based on PRS [55,56] or Polygenic Hazard Score (PHS) [57] have been proposed, rather than using a series of individual SNPs. These are based on the presence/absence of significant individual SNPs and allow to collapse all the genetic information into a single score per subject. PRS is a statistical index to estimate a subject's genetic liability to a trait or disease involving the most significant SNPs according to previous analyses, typically GWAS. Moreover, approaches based on genetic feature reduction have started to be investigated in

IG studies. In particular, by relying on a mass univariate approach, Hibar and colleagues [58], employed Principal Component Analysis (PCA) to summarize the SNPs for each gene and associated it with each brain voxel in T1-w MRI data from the ADNI first phase. While no associations survived multiple comparison adjustments, several genes known for their association with AD or brain functions were identified before correction. To further overcome the issues arising when the spatial information in images or the effect of multiple genetic variants are not taken into account in the modeling, a novel method was developed based on the random field theory and multi-locus least square kernel machines to evaluate the joint effect of multiple SNPs within each gene on more than 30 000 brain voxels [59]. The authors applied this approach to the same ADNI cohort, demonstrating this was more sensitive compared with voxel-wise single-locus approaches and identifying a number of genes as having significant associations with volumetric changes, among which GRIN2B had a prominent role. Along the same line, Le Floch et al. [30] demonstrated the importance of a pre-filtering step on individual SNPs before any multivariate analysis which can improve performance for both PLS and CCA-based methods. In [60], Wang et al. proposed a sparse multivariate multiple regression model, where SNPs were grouped by genes and the estimation of the regression coefficients was based on penalized least squares and grouping structure. More recently, Greenlaw et al. [61] extended this approach by proposing a Bayesian group sparse regression which takes into account the sparsity at the gene level. The above methods were applied on the first ADNI cohorts (sMRI data) and using a preselection of around 40 genes (and related SNPs) associated with AD in the literature.

An additional important approach in this framework is represented by the SNP set analyses which allow to improve the detection power w.r.t. individual SNP analysis, combining the effects of multiple variants together and identifying multi-locus mechanisms for complex disease. SNP sets are defined by LD blocks, genes, pathways or other criteria, which may offer biological insights for interpreting results. Different strategies have been proposed in this respect. Some methods select individual SNPs from different genomic regions associated with a given disease from literature [62] or resulting from independent analyses [55,56]. Other methods select a single gene or SNP set based on prior knowledge and examine the joint effects of multiple SNPs within this gene or set [63]. Finally, some methods exploit data-driven strategies to identify multiple SNP sets from the entire genome [23–25].

In IG, SNP set methods are used to select all SNPs belonging to significant SNP sets and then create regression models for associations with IDPs using these individual SNPs as a genetic feature. In this way, although the SNPs are selected according to a SNP set approach, the association with IDPs is made at individual SNP level. To overcome this limitation, global scores for each significant SNP set can be used as genetic features in regression models to probe the association with IDPs at SNP set level. In this work, as a novelty with respect to the previous approaches, we firstly extracted genes from the selected ADNI-3 cohort using SKAT, then we introduced a gene variant score that gives for each subject and each SKAT significant gene a measure of the extent to which all the SNPs in a given gene are mutated. Such gene variant score allows to reduce the number of variables in the multivariate model switching to the gene-wise level approach which, starting from a set of SKAT genes, takes into account the absence or presence of all SNPs in the respective gene. The gene variant score computed on the 408 significant SKAT genes were considered as genetic input for the PLS model.

Significant latent components. More in detail of the proposed PLS model, results highlighted that three PLS components, namely LC1, LC2 and LC5, span the latent space encompassing a significant separation between PAT and CN for both genotype and phenotype. Such significance was confirmed by the projection on the obtained latent space of the unseen validation cohort, holding the same class distribution of the discovery set. In fact, the separation between PAT and CN remained significant also for the validation set on both the phenotype and the genotype, the latter being particularly evident ($p < 1e^{-02}$) in the LC2. The global model significance was finally confirmed through the permutation test attaining a *p*-value of 0.001. The great advantage of the PLS model relies on its straightforward explainability. In fact, by analyzing the fitted feature weights it allows to retrieve the features driving the association between imaging and genetic features, separately for each component. A twofold result can hence be extracted: i) The intra-genotype and intra-phenotype relationships, that is observing how the different features belonging to the same data source are related to each other; ii) The analysis of the association between genotype and phenotype, highlighting those features that have a greater influence on the latent space derivation. Framing these concepts on our model allowed exploring the association between brain morphometric measures and the variant score of genes selected from the SKAT model. Multiple association patterns could be detected among the input features in the latent space. Particular attention will be given to the anticorrelations between IDPs and gene variant scores, which were indeed predominant in the LCs. The statistical analysis on the input features was instrumental to further elucidating the links between genotype and phenotype.

Relevant IDPs-gene interactions in the first latent component. Analyzing each component, in the LC1, the major role was played by the anticorrelation between the cortical thickness features (in particular supramarginal gyrus, middle/inferior/superior temporal gyri and rostral middle, superior frontal gyri) and the variant scores associated to genes RIF1, ATP6V1G2, NFASC, FBXO43, KCNA7 and KCNJ6. RIF1 associated score resulted significantly higher in PAT compared to CN, and this trend could hence be extended to all its correlated genes in this LC. Interestingly, RIF1, ATP6V1G2, NFASC and KCNJ6 were generally expressed in the brain, as a result of our transcriptomic analysis, with NFASC being particularly expressed in frontal lobe among the other regions and KCNJ6 in the hippocampus. NFASC gene is highly investigated in association with AD, transcripts were shown to be involved in synapse formation and stabilization, and were found as elevated in the subjects with MCI converting to AD compared with stable MCI as well as significantly correlated with p-tau concentration[64]. Moreover, Duits et al. concluded that, together with other peptides, NFASC transcripts could have a role in early events in the AD pathophysiological cascade. The other mentioned genes appeared also to have a role in neurodegeneration, even if their involvement in brain is still under investigation. Of interest, the RIF1 was found to protect telomers and chromosome breaks, which is in turn a process involved in brain aging [65]. Concerning ATP6V1G2, its main role appeared to be related to neurons' energy metabolism, in particular lysosome acidification. Noori et al. [66], found the ATP6V1G2 was among the genes being downregulated in neurodegenerative diseases. This downregulation may result in short ATP supply in neurons due to the failure of energy metabolism, which is however highly needed for protein clearance mechanisms. Finally, KCNA7 and KCNJ6 are part of the *Neuronal System* pathway (R-HSA-112316) and belong to the voltage-gated potassium channel gene family. In particular, KCNJ6 is associated with Down's syndrome [67], which has a well-established increased risk for AD [68]. No direct interaction between KCNA7 and AD was found, despite potassium channels are becoming a target for the treatment of neurological disorders and autoimmune diseases [69].

Relevant IDPs-gene interactions in the second latent component. Moving to LC2, this showed the most significant differences between the latent projections of PAT and CN subjects. It was mainly defined by an anticorrelation between subcortical volumes, among which hippocampus, putamen end pallidum had the highest weights, and HNMT, CACNA1A, FMO2 and EPHX1 gene variant scores. Of interest, HNMT was also among the genes showing a significant increase in the variant score for PAT compared to CN (Mann Whitney U-test, uncorrected).

In literature, it was found to be correlated with intellectual disability [70] in a cohort of patients affected by nonsyndromic autosomal recessive intellectual disability. CACNA1A, FMO2 and EPHX1 instead belonged to the four AD-related pathways highlighted in Section 3.1. More in detail, CACNA1A (*Neuronal System* pathway, R-HSA-112316) was demonstrated to be linked with familial AD in a cohort of patients presenting cerebellar damage with amyloid plaques [71]. EPHX1 (*Biological Oxidation* pathway, R-HSA-211859) has been highly investigated in literature so far. Transcripts have been detected in various areas of the brain such as cerebellum, frontal, occipital, pons, red nucleus, and substantia nigra regions. Indeed, EPHX1 was recorded as highly expressed brain-wide by the transcriptomic analysis and its role in the pathogenesis of neurodegeneration was supported by previous findings demonstrating a differential expression in patients with AD [72]. This finding well relates to our results according to which the EPHX1 variant score was anticorrelated with subcortical volumes, which also showed a significant decrease in PAT compared to CN, while it was correlated with cerebellar thickness modulations. Finally, FMO2 belongs to the *Biological Oxidation* pathway (R-HSA-211859), however its direct role in neurodegeneration has not been yet demonstrated.

Relevant IDPs-gene interactions in the fifth latent component. Finally, LC5 was mainly defined by the anticorrelation of BCAS1, GLT6D1, TMPRSS15, COL6A3 and KCNH5 with entorhinal cortex, fusiform gyrus and temporal pole thicknesses. Of note, KCNH5 showed a significant increase in its variant score for PAT compared to CN (Mann Whitney U-test, uncorrected), hence strengthening the hypothesis that the increase in such gene variant score, as well as the one of its correlated genes in LC5, is linked to a decrease in the thickness values of the mentioned brain regions. More in detail of the genes, BCAS1 is involved in the process of myelination. In fact, an explorative proteomic study of the dentate terminal zone showed that, in that region, its transcripts were among the top 10 decreased proteins showing the largest changes in AD [73]. Of interest, the dentate gyrus is part of the temporal lobe, which is indeed among the most important regions for this LC. Moreover, as a result of our transcriptomic analysis, BCAS1 also showed a brain-wide high expression. GLT6D1 was found to be associated with periodontitis. Despite its apparent distance from AD, recent experimental studies indicated that a periodontitis-causing bacterium might be a causal factor for AD since it was identified in the brain of AD patients, while in mice it provoked brain colonization and increased production of amyloid-β [74]. However, a recent bidirectional Mendelian randomization study to examine the potential causal relationship between chronic periodontitis and AD did not result in significant evidence [75]. Further studies are hence needed to deeply investigate such association. Concerning TMPRSS15, in some early-onset patients with AD induced by APP duplication (due to down syndrome), the duplicated region also contains TMPRSS15, which is hypothesized to participate in neurogenesis and/or APP metabolism [76], as detailed for gene KCNJ6, relevant in LC1. Interestingly, COL6A3 was found expressed in particular in superior temporal gyrus, which is among the significant brain regions whose thickness was decreased in PAT compared to CN. In literature, this gene has been associated with the Collagen VI protein whose lack was demonstrated to have a role in neurodegeneration [77]. Moreover, variants in this gene were found in patients affected by recessive isolated dystonia, a human brain disorder [78]. Finally, KCNH5 (*Neuronal System* pathway, R-HSA-112316) was found to be particularly expressed in entorhinal cortex, which in turn is among the most relevant regions in LC5. This gene encodes a member of voltage-gated potassium channels. Members of this family have diverse functions, including regulating neurotransmitters. It also appeared to have a role in neurodegeneration [79] even if an extensive analysis is not yet present in literature.

Overall, through the PLS model we obtained a latent representation of the input features dominated by significant genotype–phenotype associations. Most of the variant scores associated to the genes standing in the highest positions in the LCs weights were correlated with what is well known for the phenotype in AD and moreover they were found to be related to neurodegeneration as well as expressed in brain. The most relevant findings were the correlation between the EPHX1 variant score and a decrease in subcortical volumes and the correlation between the BCAS1 variant score and a significant decrease in temporal lobe thickness, as discussed in the previous paragraphs. Besides these well assessed associations, with the PLS model we retrieved multiple additional genotype–phenotype associations which are still underinvestigated in literature. Among the others the NFASC and the ATP6V1G2, were among the most relevant gene variant scores for the LC1. They are highly studied in AD but still not related with brain modulations. Our model, instead, found a significant association between the increase of the associated variant score and a decrease in temporal and frontal gyri cortical thicknesses which deserves further analysis. Moreover, in the LC5 we found multiple genes, namely GLT6D1, TMPRSS15 and COL6A3, whose involvement in AD has still not been fully proven but which resulted as significantly associated with decreased morphometric values in well known AD affected regions as the entorhinal cortex, fusiform gyrus and temporal lobe. Therefore, in summary, the PLS model allowed on one side to retrieve well assessed genotype–phenotype association whose role in the disease was already established in the current literature, and on the other side to unveil highly relevant associations between still not AD-related genes and decreased morphometric values in brain regions with a prominent role in AD, opening the way to further exploration directions.

Study limitations and future directions. We have to acknowledge some limitations in the current study. First of all we recognize the small sample size of our cohort, especially concerning patients data. This was due to the still limited number of subjects available in ADNI-3 phase (ongoing). However, this did not impact on the significance of the results thanks to the adoption of the gene variant scores. However, the use of different validation techniques such as bootstrap analysis, could not be performed. Meanwhile, ADNI-3 cohort includes the most complete set of imaging acquisition in ADNI database hence it will allow the inclusion of different IDPs providing different views on the brain modulations. This approach would be of high interest being the AD an intrinsically multifaceted disease. Along this line, the present study based on T1-w MRI could be considered as benchmark and starting point for future analysis. Diffusion MRI and functional MRI derived IDPs, such as tract-based measures or connectivity features could be included in order to investigate how both the microstructure and function are affected in the AD continuum and associated with gene variants. The gene variant score introduced in this work was computed on all SNPs located in the same gene, with each SNP being equally weighted. It could be interesting to modify the gene variant score in order to weight the SNPs based, for example, on the associations of individual SNPs with the disease or imaging phenotype (i.e. p-value from GWAS). In this direction, more sophisticated, still explainable or interpretable models could be introduced in order to account for the multi-channel information which cannot be successfully addressed through the classical PLS model which allows only the inclusion of two channels, while keeping a clear interpretation of the results, in the input matrices X and Y [39].

5. Conclusions

The presented study provides evidence of a joint variation between grey matter atrophy and gene variant scores in AD, relying on an explainable multivariate model. These associations described a latent representation of the input features that is dominated by significant genotype–phenotype associations, which have been further validated also through the transcriptomic analysis. This approach allowed uncovering previously established as well as new gene–phenotype associations, shedding new light on the underlying mechanisms of neurodegeneration in AD continuum. By focusing on genes expressed in the brain and known to be involved in the disease, we were able to detect significant correlations with AD phenotype between the BCAS1 and EPHX1 derived variant scores with decreased subcortical volumes and temporal lobe thickness respectively, confirming the importance of these genes in AD. Additionally, our analysis identified previously unexplored gene–phenotype associations, such as NFASC and ATP6V1G2, highly studied for AD, whose variant scores were found to be significantly associated with decreased cortical thickness in temporal and frontal gyri. Finally, we discovered new genes such as GLT6D1, TMPRSS15, and COL6A3, whose involvement in AD is still not fully proven, that were however significantly associated with decreased morphometric values in well-known AD-affected regions. This work hence proposed new ways to investigate the genotype–phenotype interactions in a restricted study cohort highlighting that simple yet explainable models can still allow uncovering associations that are descriptive of the underlying mechanisms of neurodegeneration in AD continuum. By identifying key genotype–phenotype interactions, new insights into the underlying mechanisms of the disease were provided, opening up new avenues for future research.

CRediT authorship contribution statement

Federica Cruciani: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft. **Antonino Aparo:** Data curation, Investigation, Software, Writing – original draft. **Lorenza Brusini:** Data curation, Investigation, Writing – review & editing. **Carlo Combi:** Writing – review & editing. **Silvia F. Storti:** Writing – review & editing. **Rosalba Giugno:** Conceptualization, Supervision, Writing – review & editing. **Gloria Menegaz:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Ilaria Boscolo Galazzo:** Conceptualization, Investigation, Supervision, Writing – original draft.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gloria Menegaz reports that financial support was provided by CariVerona Foundation and by the Italian Ministry of Education, University and Research.

Acknowledgments

This work was partly supported by Fondazione CariVerona (Bando Ricerca Scientifica di Eccellenza 2018, EDIPO project, num. 2018.0855.2019) and MIUR D.M. 737/2021 "AI4Health: empowering neurosciences with eXplainable AI methods". Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbi.2023.104569.

References

[1] Li Shen, Paul M. Thompson, Brain imaging genomics: integrated analysis and machine learning, Proc. IEEE 108 (1) (2019) 125–162.

[2] Jingyu Liu, Vince D. Calhoun, A review of multivariate analyses in imaging genetics, Front. Neuroinformatics 8 (2014) 29.

[3] Harald Hampel, Rhoda Au, Soeren Mattke, Wiesje M. van der Flier, Paul Aisen, Liana Apostolova, Christopher Chen, Min Cho, Susan De Santi, Peng Gao, et al., Designing the next-generation clinical care pathway for Alzheimer's disease, Nat. Aging 2 (8) (2022) 692–703.

[4] Emma Nichols, Jaimie D. Steinmetz, Stein Emil Vollset, Kai Fukutaki, Julian Chalek, Foad Abd-Allah, Amir Abdoli, Ahmed Abualhasan, Eman Abu-Gharbieh, Tayyaba Tayyaba Akram, et al., Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019, Lancet Public Health 7 (2) (2022) e105–e125.

[5] Clifford R. Jack Jr., David A. Bennett, Kaj Blennow, Maria C. Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M. Holtzman, William Jagust, Frank Jessen, Jason Karlawish, et al., NIA-AA research framework: toward a biological definition of Alzheimer's disease, Alzheimer's Dementia 14 (4) (2018) 535–562.

[6] Giovanni B. Frisoni, Nick C. Fox, Clifford R. Jack, Philip Scheltens, Paul M. Thompson, The clinical use of structural MRI in Alzheimer disease, Nat. Rev. Neurol. 6 (2) (2010) 67–77.

[7] Giovanni B. Frisoni, Daniele Altomare, Dietmar Rudolf Thal, Federica Ribaldi, Rik van der Kant, Rik Ossenkoppele, Kaj Blennow, Jeffrey Cummings, Cornelia van Duijn, Peter M. Nilsson, et al., The probabilistic model of Alzheimer disease: the amyloid hypothesis revised, Nat. Rev. Neurosci. 23 (1) (2022) 53–66.

[8] Meredith N. Braskie, Paul M. Thompson, Understanding cognitive deficits in Alzheimer's disease based on neuroimaging findings, Trends Cogn. Sci. 17 (10) (2013) 510–516.

[9] Lorenzo Pini, Michela Pievani, Martina Bocchetta, Daniele Altomare, Paolo Bosco, Enrica Cavedo, Samantha Galluzzi, Moira Marizzoni, Giovanni B. Frisoni, Brain atrophy in Alzheimer's disease and aging, Ageing Res. Rev. 30 (2016) 25–48.

[10] Jean-Charles Lambert, Carla A. Ibrahim-Verbaas, Denise Harold, Adam C. Naj, Rebecca Sims, Céline Bellenguez, Gyungah Jun, Anita L. DeStefano, Joshua C. Bis, Gary W. Beecham, et al., Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease, Nature Genet. 45 (12) (2013) 1452–1458.

[11] Brian W. Kunkle, Benjamin Grenier-Boley, Rebecca Sims, Joshua C. Bis, Vincent Damotte, Adam C. Naj, Anne Boland, Maria Vronskaya, Sven J. Van Der Lee, Alexandre Amlie-Wolf, et al., Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing, Nature Genet. 51 (3) (2019) 414–430.

[12] Iris E. Jansen, Jeanne E. Savage, Kyoko Watanabe, Julien Bryois, Dylan M. Williams, Stacy Steinberg, Julia Sealock, Ida K. Karlsson, Sara Hägg, Lavinia Athanasiu, et al., Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk, Nature Genet. 51 (3) (2019) 404–413.

[13] Yu Xin, Jinhua Sheng, Miao Miao, Luyun Wang, Ze Yang, He Huang, A review of imaging genetics in Alzheimer's disease, J. Clin. Neurosci. 100 (2022) 155–163.

[14] Susanne G. Mueller, Michael W. Weiner, Leon J. Thal, Ronald C. Petersen, Clifford Jack, William Jagust, John Q. Trojanowski, Arthur W. Toga, Laurel Beckett, The Alzheimer's disease neuroimaging initiative, Neuroimaging Clin. 15 (4) (2005) 869–877.

[15] Michael W. Weiner, Dallas P. Veitch, Paul S. Aisen, Laurel A. Beckett, Nigel J. Cairns, Robert C. Green, Danielle Harvey, Clifford R. Jack Jr., William Jagust, John C. Morris, et al., The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement, Alzheimer's Dementia 13 (5) (2017) 561–571.

[16] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al., UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age, PLoS Med. 12 (3) (2015) e1001779.

[17] Paul M. Thompson, Jason L. Stein, Sarah E. Medland, Derrek P. Hibar, Alejandro Arias Vasquez, Miguel E. Renteria, Roberto Toro, Neda Jahanshad, Gunter Schumann, Barbara Franke, et al., The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data, Brain Imaging Behav. 8 (2) (2014) 153–182.

[18] Marco Lorenzi, Andre Altmann, Boris Gutman, et al., Susceptibility of brain atrophy to TRIB3 in Alzheimer's disease, evidence from functional prioritization in imaging genetics, Proc. Natl. Acad. Sci. 115 (12) (2018) 3162–3167.

[19] Natalia Vilor-Tejedor, Diego Garrido-Martín, Blanca Rodriguez-Fernandez, Sander Lamballais, Roderic Guigó, Juan Domingo Gispert, Multivariate analysis and modelling of multiple brain endophenotypes: Let's MAMBO! Comput. Struct. Biotechnol. J. 19 (2021) 5800–5810.

[20] Yosef Hochberg, Multiple Comparison Procedures, in: Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 1987.

[21] Chun Chieh Fan, Robert Loughnan, Carolina Makowski, Diliana Pecheva, Chi-Hua Chen, Donald J. Hagler, Wesley K. Thompson, Nadine Parker, Dennis van der Meer, Oleksandr Frei, et al., Multivariate genome-wide association study on tissue-sensitive diffusion metrics highlights pathways that shape the human brain, Nat. Commun. 13 (1) (2022) 1–10.

[22] Andre Altmann, Marzia A. Scelsi, Maryam Shoai, Eric de Silva, Leon M. Aksman, David M. Cash, John Hardy, Jonathan M Schott, Alzheimer's Disease Neuroimaging Initiative, A comprehensive analysis of methods for assessing polygenic burden on Alzheimer's disease pathology and risk beyond APOE, Brain Commun. 2 (1) (2020) fcz047.

[23] Michael C. Wu, Seunggeun Lee, Tianxi Cai, Yun Li, Michael Boehnke, Xihong Lin, Rare-variant association testing for sequencing data with the sequence kernel association test, Am. J. Hum. Genet. 89 (1) (2011) 82–93.

[24] Priyanka Nakka, Benjamin J. Raphael, Sohini Ramachandran, Gene and network analysis of common variants reveals novel associations in multiple complex diseases, Genetics 204 (2) (2016) 783–798.

[25] Zhao-Hua Lu, Hongtu Zhu, Rebecca C. Knickmeyer, Patrick F. Sullivan, Stephanie N. Williams, Fei Zou, Alzheimer's Disease Neuroimaging Initiative, Multiple SNP set analysis for genome-wide association studies through Bayesian latent variable selection, Genet. Epidemiol. 39 (8) (2015) 664–677.

[26] Kwangsik Nho, Sungeun Kim, Emrin Horgusluoglu, Shannon L. Risacher, Li Shen, Dokyoon Kim, Seunggeun Lee, Tatiana Foroud, Leslie M. Shaw, John Q. Trojanowski, et al., Association analysis of rare variants near the APOE region with CSF and neuroimaging biomarkers of Alzheimer's disease, BMC Med. Genom. 10 (1) (2017) 45–52.

[27] Joshua C. Bis, Xueqiu Jian, Brian W. Kunkle, Yuning Chen, Kara L. Hamilton-Nelson, William S. Bush, William J. Salerno, Daniel Lancour, Yiyi Ma, Alan E. Renton, et al., Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation, Mol. Psychiatry 25 (8) (2020) 1859–1875.

[28] Pascal Zille, Vince D. Calhoun, Yu-Ping Wang, Enforcing co-expression within a brain-imaging genomics regression framework, IEEE Trans. Med. Imaging 37 (12) (2017) 2561–2571.

[29] Xiaoke Hao, Chanxiu Li, Lei Du, Xiaohui Yao, Jingwen Yan, Shannon L. Risacher, Andrew J. Saykin, Li Shen, Daoqiang Zhang, Mining outcome-relevant brain imaging genetic associations via three-way sparse canonical correlation analysis in Alzheimer's disease, Sci. Rep. 7 (1) (2017) 1–12.

[30] Édith Le Floch, Vincent Guillemot, Vincent Frouin, Philippe Pinel, Christophe Lalanne, Laura Trinchera, Arthur Tenenhaus, Antonio Moreno, Monica Zilbovicius, Thomas Bourgeron, et al., Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse Partial Least Squares, Neuroimage 63 (1) (2012) 11–24.

[31] Claudia Grellmann, Sebastian Bitzer, Jane Neumann, Lars T. Westlye, Ole A. Andreassen, Arno Villringer, Annette Horstmann, Comparison of variants of canonical correlation analysis and partial least squares for combined analysis of MRI and genetic data, Neuroimage 107 (2015) 289–310.

[32] Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, Stephen M. Smith, Fsl, Neuroimage 62 (2) (2012) 782–790.

[33] Bruce Fischl, David H. Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, et al., Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain, Neuron 33 (3) (2002) 341–355.

[34] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A.R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I.W. De Bakker, Mark J. Daly, et al., PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (3) (2007) 559–575.

[35] Daniel S. Himmelstein, Sergio E. Baranzini, Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes, PLoS Comput. Biol. 11 (7) (2015) e1004259.

[36] Marc Gillespie, Bijay Jassal, Ralf Stephan, Marija Milacic, Karen Rothfels, Andrea Senff-Ribeiro, Johannes Griss, Cristoffer Sevilla, Lisa Matthews, Chuqiao Gong, et al., The reactome pathway knowledgebase 2022, Nucleic Acids Res. 50 (D1) (2022) D687–D692.

[37] Heba Elshatoury, Federica Cruciani, Francesco Zumerle, Silvia F. Storti, André Altmann, Marco Lorenzi, Gholamreza Anbarjafari, Gloria Menegaz, Ilaria Boscolo Galazzo, Disentangling the association between genetics and functional connectivity in Mild Cognitive Impairment, in: 2021 IEEE EMBS BHI, 2021, pp. 1–4.

[38] Federica Cruciani, André Altmann, Marco Lorenzi, Gloria Menegaz, Ilaria Boscolo Galazzo, What PLS can still do for imaging genetics in Alzheimer's disease, in: 2022 IEEE EMBS BHI, 2022, pp. 1–4.

[39] Herman Wold, Nonlinear iterative partial least squares (NIPALS) modelling: some current developments, in: Multivariate Analysis–III, Elsevier, 1973, pp. 383–407.

[40] Anthony Randal McIntosh, Nancy J. Lobaugh, Partial least squares analysis of neuroimaging data: applications and advances, Neuroimage 23 (2004) S250–S263.

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[42] Evelina Sjöstedt, Wen Zhong, Linn Fagerberg, Max Karlsson, Nicholas Mitsios, Csaba Adori, Per Oksvold, Fredrik Edfors, Agnieszka Limiszewska, Feria Hikmet, et al., An atlas of the protein-coding genes in the human, pig, and mouse brain, Science 367 (6482) (2020) eaay5947.

[43] Vivian Tam, Nikunj Patel, Michelle Turcotte, Yohan Bossé, Guillaume Paré, David Meyre, Benefits and limitations of genome-wide association studies, Nature Rev. Genet. 20 (8) (2019) 467–484.

[44] Paula I. Moreira, Akihiko Nunomura, Masao Nakamura, Atsushi Takeda, Justin C. Shenk, Gjumrakch Aliev, Mark A. Smith, George Perry, Nucleic acid oxidation in Alzheimer disease, Free Radic. Biol. Med. 44 (8) (2008) 1493–1505.

[45] Eun-Joo Shin, Chu Xuan Duong, Xuan-Khanh Thi Nguyen, Zhengyi Li, Guoying Bing, Jae-Hyung Bach, Dae Hun Park, Keiichi Nakayama, Syed F. Ali, Anumantha G. Kanthasamy, et al., Role of oxidative stress in methamphetamine-induced dopaminergic toxicity mediated by protein kinase Cδ, Behav. Brain Res. 232 (1) (2012) 98–113.

[46] Constance Chace, Deborah Pang, Catherine Weng, Alexis Temkin, Simon Lax, Wayne Silverman, Warren Zigman, Michel Ferin, Joseph H. Lee, Benjamin Tycko, et al., Variants in CYP17 and CYP19 cytochrome P450 genes are associated with onset of Alzheimer's disease in women with down syndrome, J. Alzheimer's Dis. 28 (3) (2012) 601–612.

[47] Natalia Mast, Aicha Saadane, Ana Valencia-Olvera, James Constans, Erin Maxfield, Hiroyuki Arakawa, Young Li, Gary Landreth, Irina A. Pikuleva, Cholesterol-metabolizing enzyme cytochrome P450 46A1 as a pharmacologic target for Alzheimer's disease, Neuropharmacology 123 (2017) 465–476.

[48] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, et al., An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest, Neuroimage 31 (3) (2006) 968–980.

[49] Razvan Marinescu, Arman Eshaghi, Daniel Alexander, Polina Golland, Brain-Painter: A software for the visualisation of brain structures, biomarkers and associated pathological processes, 2019, arXiv preprint arXiv:1905.08627.

[50] Rui-Xian Li, Ya-Hui Ma, Lan Tan, Jin-Tai Yu, Prospective biomarkers of Alzheimer's disease: A systematic review and meta-analysis, Ageing Res. Rev. (2022) 101699.

[51] Adrià Casamitjana, Paula Petrone, José Luis Molinuevo, et al., Projection to latent spaces disentangles pathological effects on brain morphology in the asymptomatic phase of Alzheimer's disease, Front. Neurology 11 (2020) 648.

[52] Aurina Arnatkeviciute, Ben D. Fulcher, Mark A. Bellgrove, Alex Fornito, Imaging transcriptomics of brain disorders, Biol. Psychiatry Glob. Open Sci. (2021).

[53] Xiaofeng Zhu, Heung-Il Suk, Heng Huang, Dinggang Shen, Structured sparse low-rank regression model for brain-wide and genome-wide associations, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 344–352.

[54] Xiaoqian Wang, Hong Chen, Jingwen Yan, Kwangsik Nho, Shannon L. Risacher, Andrew J Saykin, Li Shen, Heng Huang, ADNI, Quantitative trait loci identification for brain endophenotypes via new additive model with random networks, Bioinformatics 34 (17) (2018) i866–i874.

[55] Shaun M. Purcell, Naomi R. Wray, Jennifer L. Stone, Peter M. Visscher, Michael C. O'Donovan, Patrick F. Sullivan, Pamela Sklar, Shaun M. Purcell, Jennifer L. Stone, et al., Common polygenic variation contributes to risk of schizophrenia and bipolar disorder, Nature 460 (7256) (2009) 748–752.

[56] Valentina Escott-Price, Rebecca Sims, Christian Bannister, Denise Harold, Maria Vronskaya, Elisa Majounie, Nandini Badarinarayan, Gerad/Perades, IGAP consortia, Kevin Morgan, et al., Common polygenic variation enhances risk prediction for Alzheimer's disease, Brain 138 (12) (2015) 3673–3684.

[57] Rahul S. Desikan, Chun Chieh Fan, Yunpeng Wang, Andrew J. Schork, Howard J. Cabral, L. Adrienne Cupples, Wesley K. Thompson, Lilah Besser, Walter A. Kukull, Dominic Holland, et al., Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score, PLoS Med. 14 (3) (2017) e1002258.

[58] Derrek P. Hibar, Jason L. Stein, Omid Kohannim, Neda Jahanshad, Andrew J. Saykin, Li Shen, Sungeun Kim, Nathan Pankratz, Tatiana Foroud, Matthew J. Huentelman, et al., Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects, Neuroimage 56 (4) (2011) 1875–1891.

[59] Tian Ge, Jianfeng Feng, Derrek P. Hibar, Paul M. Thompson, Thomas E. Nichols, Increasing power for voxel-wise genome-wide association studies: the random field theory, least square kernel machines and fast permutation procedures, Neuroimage 63 (2) (2012) 858–873.

[60] Hua Wang, Feiping Nie, Heng Huang, Sungeun Kim, Kwangsik Nho, Shannon L Risacher, Andrew J. Saykin, Li Shen, Alzheimer's Disease Neuroimaging Initiative, Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort, Bioinformatics 28 (2) (2012) 229–237.

[61] Keelin Greenlaw, Elena Szefer, Jinko Graham, Mary Lesperance, Farouk S. Nathoo, Alzheimer's Disease Neuroimaging Initiative, A Bayesian group sparse multi-task regression model for imaging genetics, Bioinformatics 33 (16) (2017) 2513–2522.

[62] Liana G. Apostolova, Shannon L. Risacher, Tugce Duran, Eddie C. Stage, Naira Goukasian, John D. West, Triet M. Do, Jonathan Grotts, Holly Wilhalme, Kwangsik Nho, et al., Associations of the top 20 Alzheimer disease risk variants with brain amyloidosis, JAMA Neurology 75 (3) (2018) 328–341.

[63] Zhiyuan Xu, Chong Wu, Wei Pan, Alzheimer's Disease Neuroimaging Initiative, et al., Imaging-wide association study: integrating imaging endophenotypes in GWAS, Neuroimage 159 (2017) 159–169.

[64] Flora H. Duits, Gunnar Brinkmalm, Charlotte E. Teunissen, Ann Brinkmalm, Philip Scheltens, Wiesje M. Van der Flier, Henrik Zetterberg, Kaj Blennow, Synaptic proteins in CSF as potential novel biomarkers for prognosis in prodromal Alzheimer's disease, Alzheimer's Res. Ther. 10 (1) (2018) 1–9.

[65] Gabriele A. Fontana, Julia K. Reinert, Nicolas H. Thomä, Ulrich Rass, Shepherding DNA ends: Rif1 protects telomeres and chromosome breaks, Microb. Cell 5 (7) (2018) 327.

[66] Ayush Noori, Aziz M. Mezlini, Bradley T. Hyman, Alberto Serrano-Pozo, Sudeshna Das, Systematic review and meta-analysis of human transcriptomics reveals neuroinflammation, deficient energy metabolism, and proteostasis failure across neurodegeneration, Neurobiol. Dis. 149 (2021) 105225.

[67] Alexander M. Kleschevnikov, Jessica Yu, Jeesun Kim, Larisa V. Lysenko, Zheng Zeng, Y. Eugene Yu, William C. Mobley, Evidence that increased Kcnj6 gene dose is necessary for deficits in behavior and dentate gyrus synaptic plasticity in the Ts65Dn mouse model of down syndrome, Neurobiol. Dis. 103 (2017) 1–10.

[68] Ira T. Lott, Elizabeth Head, Dementia in Down syndrome: unique insights for Alzheimer disease research, Nat. Rev. Neurol. 15 (3) (2019) 135–147.

[69] Heike Wulff, Boris S. Zhorov, K+ channel modulators for the treatment of neurological disorders and autoimmune diseases, Chem. Rev. 108 (5) (2008) 1744–1773.

[70] Abolfazl Heidari, Chanakan Tongsook, Reza Najafipour, Luciana Musante, Nasim Vasli, Masoud Garshasbi, Hao Hu, Kirti Mittal, Amy J.M. McNaughton, Kumudesh Sritharan, et al., Mutations in the histamine N-methyltransferase gene, HNMT, are associated with nonsyndromic autosomal recessive intellectual disability, Hum. Mol. Gen. 24 (20) (2015) 5697–5710.

[71] Diego Sepulveda-Falla, Alvaro Barrera-Ocampo, Christian Hagel, Anne Korwitz, Maria Fernanda Vinueza-Veloz, Kuikui Zhou, Martijn Schonewille, Haibo Zhou, Luis Velazquez-Perez, Roberto Rodriguez-Labrada, et al., Familial Alzheimer's disease–associated presenilin-1 alters cerebellar activity and calcium homeostasis, J. Clin. Invest. 124 (4) (2014) 1552–1567.

[72] Radka Václavíková, David J. Hughes, Pavel Souček, Microsomal epoxide hydrolase 1 (EPHX1): Gene, structure, function, and role in human disease, Gene 571 (1) (2015) 1–8.

[73] Hazal Haytural, Georgios Mermelekas, Ceren Emre, Saket Milind Nigam, Steven L. Carroll, Bengt Winblad, Nenad Bogdanovic, Gaël Barthet, Ann-Charlotte Granholm, Lukas M. Orre, et al., The proteome of the dentate terminal zone of the perforant path indicates presynaptic impairment in Alzheimer disease, Mol. Cell. Proteom. 19 (1) (2020) 128–141.

[74] Vladimir Ilievski, Paulina K. Zuchowska, Stefan J. Green, Peter T. Toth, Michael E. Ragozzino, Khuong Le, Haider W. Aljewari, Neil M. O'Brien-Simpson, Eric C. Reynolds, Keiko Watanabe, Chronic oral application of a periodontal pathogen results in brain inflammation, neurodegeneration and amyloid beta production in wild type mice, PLoS One 13 (10) (2018) e0204941.

[75] Yi-Qian Sun, Rebecca C. Richmond, Yue Chen, Xiao-Mei Mai, Mixed evidence for the relationship between periodontitis and Alzheimer's disease: A bidirectional mendelian randomization study, PLoS One 15 (1) (2020) e0228206.

[76] Frances K. Wiseman, Tamara Al-Janabi, John Hardy, Annette Karmiloff-Smith, Dean Nizetic, Victor L.J. Tybulewicz, Elizabeth Fisher, André Strydom, A genetic cause of Alzheimer disease: mechanistic insights from Down syndrome, Nat. Rev. Neurosci. 16 (9) (2015) 564–574.

[77] Matilde Cescon, Peiwen Chen, Silvia Castagnaro, Ilaria Gregorio, Paolo Bonaldo, Lack of collagen VI promotes neurodegeneration by impairing autophagy and inducing apoptosis during aging, Aging (Albany NY) 8 (5) (2016) 1083.

[78] Michael Zech, Daniel D. Lam, Ludmila Francescatto, Barbara Schormair, Aaro V. Salminen, Angela Jochim, Thomas Wieland, Peter Lichtner, Annette Peters, Christian Gieger, et al., Recessive mutations in the $\alpha3$ (VI) collagen gene COL6A3 cause early-onset isolated dystonia, Am. J. Hum. Genet. 96 (6) (2015) 883–893.

[79] Richard Sherva, Alden Gross, Shubhabrata Mukherjee, Ryan Koesterer, Philippe Amouyel, Celine Bellenguez, Carole Dufouil, David A. Bennett, Lori Chibnik, Carlos Cruchaga, et al., Genome-wide association study of rate of cognitive decline in Alzheimer's disease patients identifies novel genes and pathways, Alzheimer's Dementia 16 (8) (2020) 1134–1145.