Identifying Single-Nucleotide Polymorphisms Intersecting Alzheimer's Disease Pathology and End-of-Life Traits Using Genomic Informational Field Theory (GIFT)

Sam Heysmond[1], Panagiota Kyratzi[1,2], Jonathan Wattis[3], Andras Paldi[2], Keeley J Brookes[4], Karim Kreft[5,6] , Beili Shao[5,6] and, Cyril Rauch[1,*]

[1]School of Veterinary Medicine and Science, University of Nottingham, College Road, Sutton Bonington, LE12 5RD, UK.
[2]École Pratique des Hautes Études, PSL Research University, St-Antoine Research Center, Inserm U938, 34 rue Crozatier, 75012 Paris, France.
[3]School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK.
[4]Biosciences, Clifton Campus Nottingham Trent University, Nottingham, NG11 8NS, UK
[5]Academic Neurology, Academic Unit of Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, Nottingham, UK
[6]Department of Neurology, Nottingham University Hospitals NHS Trust, Nottingham, UK

[*]**Corresponding author**: cyril.rauch@nottingham.ac.uk

Keywords: Alzheimer disease, neuropathology, end of life, lifespan, genetic path, GWAS, association.

ABSTRACT

Background: Quantitative genome-wide association studies (GWAS) primarily rely on additive linear models that compare average phenotypic differences between genotype groups. While effective for detecting common variants of moderate effect in large sample sizes, such approaches inherently reduce high-resolution phenotypic data to summary statistics (group averages), potentially limiting the detection of subtle genotype–phenotype relationships. Genomic Informational Field Theory (GIFT) is a recently developed methodology that preserves the fine-grained informational structure of quantitative traits by analysing ranked phenotypic configurations rather than relying solely on mean differences.

Methods: We applied GIFT to genetic and neuropathological data from the Brains for Dementia Research cohort, a well characterised dataset of 563 individuals, and compared its performance with conventional GWAS. Principal component analysis (PCA) derived matrix was used to derive independent quantitative traits linked to from Alzheimer's disease (AD) neuropathology measures (CERAD, Thal, Braak staging), with and without inclusion of age at death. Principal component analyses were performed using GWAS and GIFT frameworks on the same filtered genotype dataset.

Results: Both GWAS and GIFT identified genome-wide significant associations (pvalue<10^{-6}) within the APOE locus (NECTIN2–TOMM40–APOE–APOC1), demonstrating concordance with established AD genetic variants. However, GIFT detected additional significant 19 SNPs beyond those identified by GWAS. Variants associated with AD pathology implicated genes involved in amyloid processing, neuronal apoptosis, synaptic function, neuroinflammation, and metabolic regulation. Notably, GIFT identified 29 loci associated with age at death–related variation that were not detected by GWAS, highlighting genes linked to lipophagy, mitochondrial quality control, sphingolipid metabolism, frailty, and aging-related processes.

Conclusions: GIFT recapitulates canonical GWAS findings while uncovering additional biologically relevant associations. By preserving the fine-grained structure of phenotypic data distributions and detecting non-random genotype segregation across ranked trait values, GIFT enables the identification of associations that remained undetected by traditional average-based GWAS approaches. These results demonstrate that rethinking analytical representation, rather than solely increasing sample size, can expand discovery potential of genetic association studies, offering a transparent and complementary framework for quantitative genomics in deeply phenotyped datasets.

INTRODUCTION

Traditional Genome-Wide Association Studies (GWAS) assume an additive inheritance model with a binary trait or diagnosis in a logistic regression model to compare the average risk expressed as an odds ratio between genotypes of each single nucleotide polymorphism. Analysis for quantitative traits uses the sample principle with a linear regression, comparing the averages of the measured trait between genotype groups, but also provides increased statistical power to detect SNP associations for the same underlying genetic effect, due to the greater information about the variability between individuals. Therefore, utilising measures of underlying pathology of diseases might uncover more of the genetic aetiology. However, the use of averages in GWAS to compare genotypic groups inherently restricts information on individual variance and is primarily designed to detect additive linear genetic effects. Given the small effect sizes typically associated with genetic variants influencing complex traits or diseases, GWAS may fail to capture subtler effects.

Recently, a new methodology for identifying SNPs associated with traits has emerged. Known as Genomic Informational Field Theory (GIFT) and inspired by field theory in physics, this approach preserves the informational content of datasets through an alternative representation and analysis of datasets.

Traditional statistics is underpinned by frequency- or count-based visualizations, such as bar charts or histograms, which group raw data into discrete bins. When these discrete representations are interpolated in the continuum limit and conceptualized as continuous distributions, they are referred to as distribution density functions (DDFs). A key example of DDF is the normal distribution, widely employed in GWAS. Ultimately, DDFs interpreted within the frequentist probability framework provide the foundation for calculating key population-level statistical summaries (e.g., averages, variances, and measures of significance) that subsequently underpin population-level inferences. However, while DDF-based statistics and related statistical summaries reveal population-level properties, they do so at a cost: by transforming raw data into aggregated representations, the process sacrifices information about the highly resolved structure and fine-grained diversity of individual datapoints. To illustrate this point, consider a population in which phenotypic measurements are captured with high precision, rendering each data point unique, a scenario increasingly frequent with the advent of advanced measurement technologies [1]. Once these data are grouped into bins to construct frequency- or count-based plots, the individuality of each measurement is effectively erased, as all points within a bin become indistinguishable. In effect, data binning results in a substantial loss of the information originally present in the dataset's unaggregated form. Consequently, although the statistical summaries derived from DDFs may appear informative, they are ultimately products of information loss rooted in an impoverished representation of the original dataset, which imposes limits on the knowledge that can be extracted from the data. Therefore, GIFT treats each quantitative trait measurement as a unique data point and analyses it according to its rank order within the dataset, preserving individual-level information and enabling more nuanced insights. GIFT then determines the association between genotype and phenotype by computing the cumulative sum of genotypes corresponding to the ranked trait values, from small to large trait values. The cumulative sum of genotypes reported to the position in the rank defines a curve termed 'genetic path', and noted θ thereafter, captures not only the effects linked to genotypes averages (allele substitution) but also the effects linked their variances [2-5]. The significance of the association is then determined by comparing the maximal amplitude difference of the observed genetic path, θ, with that of the resulting average null-path, noted θ_0, generated using the same method but with the ranked trait values and corresponding genotypes randomly permuted an infinite number of times. Formalised theoretically [3-5], GIFT's ability to detect genotype–phenotype associations has been evaluated using simulations [3], and real-world datasets [2]. These studies demonstrate its capacity to detect small effect sizes without requiring sample sizes in the thousands.

The aim of this study was to investigate the effectiveness of PCA analysis from ordinal scale pathology categories and co-variates to increase the range of unique datapoints and therefore information to be used in GWAS and GIFT, and subsequently compare the genetic associations identified from both analyses. This investigation utilised genetic data from the Brains for Dementia Research (BDR) cohort [6], including detailed neuropathological assessments and age at death (n = 563). The study compared genetic association findings derived from conventional genome-wide association analyses (GWAS) and GIFT across two related datasets. The first dataset comprised variables specifically reflecting Alzheimer's disease (AD) neuropathology, while the second incorporated both AD neuropathological measures and age at death. This design aimed to determine whether genetic variants associated with AD pathology also contribute to variation in lifespan, or whether distinct genetic factors influence survival independently of AD-related pathological processes. We demonstrate that only GIFT identified genes *(NECTIN2*, *TOMM40*, and *APOE*) were jointly

associated with both AD neuropathology and age at death. In contrast, a substantial number of additional loci were associated with either AD pathology or age at death independently, but not with both traits simultaneously.

MATERIALS AND METHODS:

Dataset: The genotyping and quality control on the BDR cohort is detailed elsewhere [7]. This dataset comprised 563 BDR participants with complete neuropathological assessments for Alzheimer's disease (AD), including CERAD score, Thal phase, and Braak tau stage, which together form the ABC score for AD neuropathological change. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) score is a semi-quantitative amyloid rating scale (0–3) used to classify the density of neuritic β-amyloid plaques in the brain. Thal amyloid staging (0–5) characterises the hierarchical anatomical spread of β-amyloid deposition. Stage 0 indicates no amyloid deposits. Stage 1 reflects deposits confined to the neocortex, stage 2 indicates spread to allocortical regions (e.g., hippocampus), stage 3 reflects involvement of the thalamus and striatum, stage 4 indicates extension to the brainstem, and stage 5 denotes amyloid deposition in the cerebellum in addition to all previously affected regions. Braak tau staging (0–VI) categorises disease progression based on the distribution of neurofibrillary tangles (NFTs). Stage 0 indicates no NFTs. Stage I involves the medial temporal lobe. Stage II reflects spread to the entorhinal cortex and hippocampal CA1 region. Stage III involves the hippocampus, amygdala, and early neocortical regions. Stage IV indicates further spread to temporal, parietal, and occipital cortices. Stage V reflects widespread involvement of association neocortical areas, and Stage VI denotes NFTs in primary motor and sensory cortices, indicating extensive cortical involvement. Stages I–III are typically associated with preclinical AD (no significant clinical symptoms), Stages III–IV with prodromal disease often diagnosed as mild cognitive impairment, and Stages V–VI with clinically manifest dementia. These three neuropathological measures (CERAD, Thal phase, and Braak stage) are combined into the ABC score to determine the overall level of Alzheimer's disease neuropathological change. ABC classifications of "Intermediate" or "High" are considered sufficient to account for the observed dementia [8].

Generation of Quantitative Phenotypes: The ABC score for AD and, the ABC score for AD with or without age at death, were entered into a principal component analysis (PCA) to produce to set of quantitative independent variables for analysis using samples with sufficient genotype information (missingness $\leq 10\%$). Principal components with or without age at death were entered into traditional GWAS quantitative trait association analyses and GIFT to detect SNPs related to AD.

Association Analysis: Typical GWAS methodologies were employed alongside GIFT to perform association analysis. Genetic data were filtered to include individuals of interest in the study; we removed samples with no corresponding phenotypic information for Thal, CERAD and Braak Tangle as well as for missing pathology group information. The genetic data were also filtered for biallelic SNPs, selecting minor allele frequency ($MAF \geq 5\%$), missingness per SNP $\leq 10\%$ and HWE $\leq 10^{-6}$. Genetic data were aligned to the hg19 reference genome, stored at UCSC genome browser database [9], and rsIDs were imported from the latest dbSNP data for the same hg19 reference [10]. This filtering and formatting was accomplished using bcftools, vcftools and plink software (www.cog-genomics.org/plink/2.0/) [11-13]. GWAS calculations were performed using GEMMA software [14], Kinship matrices were created from the genotype data to account for population structure and relationships and added as input for GEMMA when calculating GWAS pvalues

GIFT: The filtered genetic data were passed into the GIFT algorithm where genotypes were first transformed as follow, (0|0) →(-1), (0|1 and 1|0)→(0) and (1|1)→(-1), and subsequently positioned alongside ordered phenotypic values (smallest to largest), then the cumulative sum of genotypes were computed and represented under the form of a genetic path whose amplitude was used to compute the level of significance between the genotype and the phenotype as detailed elsewhere [4]. The GIFT methodology will be introduced in the main text.

Significance threshold for GWAS and GIFT: To calculate Bonferroni thresholds for the GWAS and GIFT results, plink 2.0 [11] was used to clump the association data using the filtered genotype data as a reference to keep only the most significant and non-LD SNPs. The filters used were clump-r2 of 0.1, clump-kb of 250, and clump-p1 of 0.05. The total number of index SNPs were used to calculate 5% Bonferroni threshold as standard ($-\log_{10}(0.05/N_{ind\text{-}SNPs})$) per analyses.

Computation of significance for GIFT: As detailed elsewhere [4], for each SNP the maximal amplitudes resulting from the cumulative sum of genotypes ranked as a function of phenotypic values were used to compute the level of significance of associations.

Software and codes: All figures were designed in R software, using ggplot package and igraph software for visualisation [15]. All scripts are run on the University of Nottinghams Ada HPC cluster service utilising the slurm environment and conda environments created under conda version 23.7.4 , bioconda and conda forge [16, 17]. For full details on scripts and software packages with versions involved in analysis, please see Github repository at: https://github.com/SamHeysmond/gift_pipeline_HPC.

RESULTS

Principles underscoring GIFT in genotype-phenotype associations

GIFT is an optimal methodology for identifying associations between datasets without relying on predefined categories or data grouping. To illustrate its approach, consider a continuously varying trait measured with sufficient precision such that no two phenotypic values are identical. In biallelic organisms, three genotypic categories, bB/Bb, and BB, are present, following Mendelian inheritance. Assuming the trait is linked to these genotypes, potential associations can be explored using count or frequency plots, as shown in Figs. 1A and 1B. These plots typically reveal that each genotype occurs predominantly within specific ranges of the continuous trait, a pattern that can be confirmed using statistical tests comparing the mean trait values, as shown in Fig.1C.

Nevertheless, this approach may mask subtle associations that GIFT is specifically designed to reveal. One major drawback of categorising continuous data is the loss of detailed information, as differences within each category cannot be captured. GIFT addresses this issue by examining associations without grouping the data, preserving its full resolution and allowing for more accurate inferences, even with smaller sample sizes.

As shown in Figures 1D and 1E, if we were to know the exact value of datapoints, the absence of categories compels us to concentrate on barcodes to provide means to extract inferences. Importantly, coarse graining the barcodes recover the traditional way of representing datasets under the form of histograms. Now, concentrating on the different colours of the bars in Fig.1E, what is visible is that the distribution of colours is not uniform but that bars are segregated as a function of their colour, reflecting the positioning of the different distributions in Fig.1B. Importantly, while each bar is associated with a specific trait value in Fig.1E, this specific trait value depends on the subpopulation used and a different subpopulation would have provided a slightly different trait value for each bar. That is to say that what matters to determine inferences when considering the coloured barcode in Fig.1E is not the exact positioning of bars, but their relative positioning.

To determine whether the segregation of coloured bars has any significance, GIFT proceeds initially by numbering the different colours using +1 for red bars (genotype bb), 0 for gray bars (genotypes bB/Bb) and –1 for blue bars (genotype BB), thereby creating an initial string of numbers only containing +1, 0 and –1 as they appear from low to high trait values. In a second step, GIFT creates another string where the component, say at position i, corresponds to the cumulative sum of numbers located between the positions 1 to i from the initial string. The component of this new string for the position i is noted $\theta(i)$ and call genetic path. Then $\theta(i)$ is plotted as a function of the position i as shown in Fig.1F. The curved aspect of $\theta(i)$ results the presence of red bars (+1 bars) initially followed by grey bars (0 bars) and finally blue bars (-1 bars) as shown in Fig.1E.

This curve is then compared to another function resulting from a scenario corresponding to the null hypothesis. This scenario is obtained by randomly permutating the number +1, 0 and –1 in the initial string. We shall call this randomly permutated string a null string noted $\theta_0(i)$. However, as there are many random permutations possible of the initial string one may determine the resulting average null string after considering an infinite number of permutations. In this context, it is possible to demonstrate that the resulting average null string leads, after considering the average cumulative sum of components, to a straight-line noted $\overline{\theta}_0(i)$ (see Fig.1F), whose equation is given by, $\overline{\theta}_0(i) = \frac{N_+ - N_-}{N} i$, where N_+, N_- are the number of +1, -1 bars, and N is total number of bars in the initial string, respectively. Note that the horizontal bar on the top of θ_0 signifies that the average has been considered.

The notion of significance with GIFT is then determined by comparing the difference, $\theta(i) - \overline{\theta}_0(i)$ (Fig.1F). There are different ways to determine a measure of significance for the difference $\theta(i) - \overline{\theta}_0(i)$, which ultimately depend on theoretic considerations such as symmetries and invariances of the dataset and scientific problem studied. However, one generic way to determine a measure of significance is to recall that the segregation of microstates is primordial and that it leads to large amplitudes for $\theta(i)$. One may then concentrate on the difference between the max and min values of, $\theta(i) - \overline{\theta}_0(i)$ and determine the level of significance using hypergeometric probability distribution theory, see [4] for detailed information.

As will be demonstrated below in the context of AD, GIFT can identify associations as long as the genotypes are not randomly distributed; in other words, GIFT extracts associations that are closely linked to the means and variances of the genotypes.

Demographics of the BDR cohort

A subset of 563 samples of the BDR cohort was used in this study. Overall, the average age at death was 84.1 years (SD 8.9) and with 48.3% of the cohort being female. One hundred and seventy-four of these samples were defined as controls with no cognitive deficit and no dementia relevant pathology detected post-mortem (mean age at death = 84.7 (SD=10.0); 52.9% female). Whilst the remaining n=389 samples displayed cognitive deficits and had AD pathology confirmed at post-mortem (mean age at death = 83.8 (SD=8.3); 46.3% female), no significant differences ($p>0.05$) were observed for these characteristics between the two groups.

Principal Component Analyses

Principal Component Analysis (PCA) is a statistical dimensionality-reduction technique that transforms correlated variables into a smaller set of uncorrelated variables called principal components. Although the number of components initially equals the number of original variables, they are ordered according to the amount of variance (information) they capture from the data. Consequently, only the components that explain the largest proportion of variance are typically retained for analysis. Figures 2A and 2B show the fraction of variance explained when using AD neuropathology features alone and when combining AD neuropathology features with age at death, respectively. In Fig. 2A, the first principal component (PC1) explains 84.6% of the variance, indicating that it is the most informative component. However, in Fig. 2B, both PC1 and PC2 must be considered, as they explain 64.1% and 24.9% of the variance, respectively, together accounting for 89% of the total variance explained, thereby capturing a level of information similar to that of PC1 in Fig. 2A.

To determine how the initial variables distribute across the principal components, biplot representations were used for each dataset. In a biplot, the original variables are depicted as vectors emanating from the origin, where each vector corresponds to one variable. The direction of each vector indicates the variable's influence on the principal components, while its length reflects the strength of this contribution. The angle between vectors conveys the correlation among variables: acute angles suggest positive correlation, obtuse angles indicate negative correlation, and orthogonal vectors imply no correlation. In this representation, Figure 2C shows that the AD neuropathology features are much more closely aligned with the horizontal axis (PC1) than with the vertical axis (PC2), thereby underpinning the strong correlations among AD neuropathology features, as expected from Fig.2A. In contrast, Figure 2D shows that the AD neuropathology features and age at death are aligned along different principal components, with AD neuropathology features associated with PC1 (horizontal axis) and age at death associated with PC2 (vertical axis), indicating that AD neuropathology features and age at death are, statistically, poorly correlated in this dataset.

In the remainder of the paper, we use the notation $PC1_{AD}$ to denote the first principal component derived from AD neuropathology variables only, and the notations $PC1_{AD+Age}$ and $PC2_{AD+Age}$ to denote the first and second principal components, respectively, derived from AD neuropathology variables together with age at death.

GWAS and GIFT Manhattan plots

Because individual data points from the two datasets are informatively projected along their principal components, their coordinates can be used as numerical inputs to define what we term PC-related trait values. Since AD neuropathology variables inform only a single principal component ($PC1_{AD}$), whereas inclusion of age at death informs two principal components ($PC1_{AD+Age}$ and $PC2_{AD+Age}$) that are weakly correlated, two complementary analytical approaches can be pursued: one focusing on the two first principal components $PC1_{AD}$ and $PC1_{AD+Age}$ that, by virtue of Fig.2, primarily capture AD neuropathology–related

variations, and another focusing on the second component $PC2_{AD+Age}$, which is more strongly associated with age at death.

Figure 3 shows the Manhattan plots obtained using GWAS and GIFT. Specifically, Figure 3A presents the Manhattan plot for AD neuropathology variables only ($PC1_{AD}$). Figure 3B shows the Manhattan plot for the first principal component derived from AD pathology and age at death variables ($PC1_{AD+Age}$), while Figure 3C displays the Manhattan plot for the second principal component ($PC2_{AD+Age}$) of these same variables. Table 1 and Table 2 provide detailed information on the set of significant SNPs determined by GWAS and GIFT, respectively. To simplify notation in the tables, "A" denotes results obtained using $PC1_{AD}$ derived from AD pathology variables only (see column "A" in the Tables); "B" denotes results obtained using $PC1_{AD+Age}$ derived from AD pathology and age at death variables; and "C" denotes results obtained using the second component $PC2_{AD+Age}$ derived from AD pathology and age at death variables.

Analysis of AD neuropathology variables only ($PC1_{AD}$) identified 10 SNPs associated with the trait all within 42kb of each other around the *APOE* locus in GWAS, and at the exclusion of rs8106922 (Chr19) all of these were picked up in the GIFT analysis of the same variable. GIFT did not identify additional markers in this region but did observe significant association with an additional 4 SNPs at the 5% Bonferroni-corrected significance threshold, in *B3GALT1* (Chr2); the *SOX7-AS1/PINX1* locus (Chr8); and 2 intergenic SNPs in Chr 2 and 4 (See Tables 1 & 2).

The analysis of the pathology variables including age-at-death yielded $PC1_{AD+Age}$ and $PC2_{AD+Age}$ for analysis correlated with neuropathological measures and age at death, respectively.

GWAS analysis identified the same 10 SNPs as in the $PC1_{AD}$ analysis from the neuropathology only, with no additional SNPs, whilst GIFT identified one additional SNP, rs8103922, in the *APOE* locus to the previous pathology only results. In addition, GIFT picked up 3 of the 4 SNPs identified in the pathology only analysis, with rs7826180 (*SOX7-AS1/PINX*) just missing the significance cut-off. Furthermore, another 16 SNPs were found to be significantly associated with PC1 in the pathology plus age analysis across chromosomes 2,4,7.10, 12, 18, 21 and 22 (Table 2).

Finally, analysis of $PC2_{AD+Age}$ derived from the AD pathological features plus age-at-death yielded only a single finding from GWAS in an intergenic region on Chr7. In contrast, GIFT identified 29 SNPs associated with this measure, highlighting multiple genes across several chromosomes (Table 2). None of these SNPs had been associated with $PC1_{AD}$ or $PC1_{AD+Age}$ measures with GIFT, expect for the 4 located in the *APOE* region (rs34342646, rs71352238, rs34404554 and rs769449), which were association in both GIFT and GWAS and across all three PCs investigated.

Genetic Path Information from GIFT

To clarify why SNPs identified as significant by GWAS were associated only with AD pathology features and not when age at death was included, we compared the representational frameworks of GWAS and GIFT for rs769449 (Chr19), the most significant SNP identified by GWAS and GIFT (Tables 1 & 2), as an example. Figure 3 illustrates the methodological differences between GWAS and GIFT.

In Fig. 4A, when $PC1_{AD+Age}$ is considered, the allelic substitution effect, based partitioning of phenotypic values by genotype as modelled by GWAS, is clearly visible. Fig. 4B shows the genetic paths in the GIFT framework, highlighting a clear deviation of $\theta(i)$ from $\bar{\theta}_0(i)$. This deviation is further visible when their difference is considered, as reflected by the large-amplitude paraboloid shape of $\Delta\theta(i)$ in Fig.4C that is used to calculate level of significance of the association. A paraboloid shape with a large amplitude (Fig. 4C) arising from ranking phenotypic values, indicates that the different genotypes are not randomly distributed across the whole range of phenotypic values but are segregated in opposite direction within specific subranges, as shown in Fig. 4A using the GWAS methodology. Colouration of points with ABC pathology ordinal scales demonstrates the clear correlation of $PC1_{AD+Age}$ with clinical pathology outcomes

When age at death ($PC2_{AD+Age}$) is considered, the allelic substitution effect, which is negligible or absent in the GWAS framework (Fig. 4D), reflects the lack of a clear distinction between the average phenotypic values across genotypes. However, the genotypes are not randomly distributed across the entire range of phenotypic values and show some segregation within specific subranges (e.g., genotype AA in Fig. 4D). The genetic paths (Fig. 4E) and their differences (Fig. 4F) capture this non-random configuration suggesting, in turn, its association with age at death, and shown by the colouration of the datapoints age measures.

Finally, to demonstrate that $PC1_{AD+Age}$ and $PC2_{AD+Age}$ specifically capture AD pathology variables and age-at-death variables, respectively, age at death was overlaid onto the genetic path $\theta(i)$ from Fig. 4B, and AD pathology features were overlaid onto the genetic path $\theta(i)$ from Fig. 4E. As shown in Fig. 4G and Fig. 4H, respectively, the features become disordered when projected onto the non-corresponding genetic path.

DISCUSSION

This study has sought to utilise GIFT for a complex human disease where discreet quantitative variables are not available. Instead, principal component analysis has been implemented to turn a categorical ordinal scale of pathology into near individualised data points for quantitative analysis. Using the resultant PCs GIFT has been directly compared to GWAS to assess its ability to identify subtle genetic effects that could be missed in the traditional analyses.

The three pathological features for AD post-mortem confirmation (Thal, CERAD and Braak) alongside age-at death were entered into principal component analysis. The pathology only analysis, yielded one key PC that capture much of the variance (84.6%); interestingly the additional of age-at-death, decrease the variance captured from PC1 to 64.1%, but still aligned strongly with the pathological features. PC2 in this analysis captured a further 24.9% and aligned strongly with age-at death, suggesting that age might be less correlated with AD pathology than previous thought. Age is considered a crucial co-variable in dementia analysis given it is the largest risk factor for the disease, doubling risk for AD every 5 years after the age of 65 [18]; yet the presence of dementia can shorten life span considerably [19].

The *APOE* locus was associated in all three PC measures analysed, although the significance of these SNPs when associated with pathological traits were greater when age included in the analyses. This result reveals indirectly GIFT's sensitivity. Indeed, it is notable from Table 2, i.e., using GIFT, that the levels of significance of SNPs can differ between the two PC1-derived traits: some SNPs are associated specifically with $PC1_{AD}$, while others are associated with $PC1_{AD+Age}$. This discrepancy is, to some extent, expected, for two reasons. The first reason is that the PCA analyses informing AD pathology were performed on two distinct datasets; consequently, the first principal components, while globally similar, are mathematically different. The second reason relates to the nature of the AD pathology dataset, which is informed by scoring systems that provide numerical trait values that are sometimes identical across individuals in the cohort. This imposes a degree of coarseness in GIFT, resulting in some loss of information, which does not happen when the AD pathology and age at death dataset is used that is richer in information. This distinction is illustrated in Table 2, where the column B, corresponding to $PC1_{AD+Age}$, contains a greater number of significant SNPs than column A, which corresponds to $PC1_{AD}$. An exception is rs7826180 on chr8, whose significance is higher for $PC1_{AD}$ (column A in Table 2) than $PC1_{AD+Age}$ (column B in Table 2). Nevertheless, the level of significance obtained for rs7826180 (pvalue = $2.04 \cdot 10^{-6}$, Table 2) remains close to the Bonferroni threshold (pvalue = $1.74 \cdot 10^{-6}$), making it a strongly suggestive SNP.

Comparison of GWAS and GIFT associated SNPs with AD pathology

Strikingly, the association analyses of the two PC1s using either GWAS and GIFT method yielded the same 10 significant SNPs at the genome-wide level on Chr19 all within a 42kb region surrounding the *APOE* gene (*NECTIN2 – TOMM40 – APOE – APOC1*) and included the rs429358 SNPs that determines *APOE epsilon*. To determine whether any of the significant rsIDs common to GWAS and GIFT had previously been associated with Alzheimer's disease, the GWAS Catalog was interrogated using "Alzheimer" as the trait term to extract relevant bibliographic references [20]. The results, reported in the last column of Table 1, demonstrate that all identified SNPs have previously been associated with Alzheimer's disease.

Interestingly, a recent GWAS combining three large sample datasets together (n=7804) explored genetic association with several Alzheimer's disease and dementia-related endophenotypes, using either binary or ordinal scales of pathology severity. Pertinent to the current study several genetic associations were found with amyloid-β plaques (defined as none, mild, moderate severe), NFT Braak staging and CERAD. Analyses with each of these three traits resulted in *APOE* rs429358 being the most significant finding with additional associations for genes around the *APOE* locus (*NECTIN2, TOMM40 and APOC1*), as found in this study with using $PC1_{AD+Age}$ [21].

It is worth noting, however, that, unlike GWAS that associated SNPs on Chr19 exclusively with AD pathology features, GIFT suggests that a subset of these SNPs is also significantly associated with age at death.

Relevance of associations with AD pathology features identified by GIFT.

In addition to the *APOE* locus, GIFT identified additional SNPs where GWAS did not.

As done above, to determine whether any of the significant rsIDs determined by GIFT outside Chr19 had previously been associated with Alzheimer's disease, the GWAS Catalog was re-interrogated using the same trait term as above [20]. None of the significant rsIDs from Table 2 were previously reported as associated with AD, which prompted a direct bibliographic search.

Interestingly, rs6734633 (Chr2; *NRXN1*), rs11721706 and rs10866283 (Chr4; *TAPT1-AS1* and *SORBS2*) and rs5758837 (Chr22; *CYB5R3*) did display nominal significance with AD-pathology confirmed diagnosis in the original GWAS of the BDR cohort [6]. Furthermore, nominal significance ($p<0.05$) for SNPs on Chr12, has been observed for rs2270238 *(KLRD1)* [22] and for rs7312352 (*PPM1H*) [23], while the Chr18 SNP rs2298574 (*CDH2*) was also previously identified as nominally significant [23,24].

Additionally, genes identified through their rsID by GIFT in our analysis have been previously implicated in AD and may provide insight into disease mechanisms. *SORBS2* has been linked to cognitive deficits, cortical amyloid-β accumulation, and an increased Aβ42/Aβ40 [25,26], and has also been reported as a genetic modifier of age at onset of AD [27]. Our analysis further highlighted *SOX7-AS1*, which was found to be downregulated in circulating exosomes from AD patients compared with healthy controls [28]. Given that *SOX7-AS1* likely negatively regulates *SOX7* expression, its downregulation may lead to elevated *SOX7* levels, consistent with enhanced neuronal apoptosis. This aligns with prior studies showing that *SOX7* overexpression promotes apoptosis, whereas *SOX7* knockdown protects neuronal cells from cell death in mouse models [29].

Other notable genes include *PINX1* previously associated with late-onset AD through meta-analysis of GWAS datasets [30], and *MET*, which encodes the hepatocyte growth factor receptor, a tyrosine kinase receptor mediating neurotrophic and neuroprotective signalling [31]. Reduced *MET* expression in the cerebral cortex and hippocampus has been observed in AD [32–34].

Immune-related genes were also highlighted. *KLRD1* is a marker of natural killer cells that has been linked to neuroinflammation in AD through its overexpression in human peripheral blood mononuclear cells [35].

Genes involved in neuronal signalling and synaptic function were also identified. *ANO4*, a calcium-dependent non-selective cation channel is also a lipid scramblase expressed in CNS and endocrine tissues associated with cognitive decline in AD [34]. *CDH2* encodes N-cadherin, a key cell adhesion molecule for synapse maintenance and neuronal migration, has been associated with amyloid-β–induced synaptic impairment in post-mortem AD brains [36,37]. Finally, *CYB5R3* a membrane-bound redox enzyme crucial for cellular energy metabolism and redox balance, has been associated with AD progression [38].

Collectively, these findings demonstrate that gene names identified by GIFT capture are implicated in multiple pathways relevant to AD pathology.

Relevance of associations with age at death identified by GIFT.

Comparison of the results presented in Tables 1 and 2 shows that GWAS identifies a single SNP from an unnamed gene associated with age-at-death variables (chr7: rs6952104; Table 1), which is not detected by GIFT. However, this SNP has been linked to genes RNASSP228 and *STK31*, and has previously been identified in GWASs of cognitive performance [39,40].

In contrast, GIFT identifies multiple significant SNPs that are not captured by GWAS. To determine whether any of the significant rsIDs identified by GIFT had previously been associated in relation to a phenotype similar to age at death, the GWAS Catalog was re-interrogated using "life span" as the trait term [20], only returning rs769449 on Chr19 and within the *APOE* region [41]. Note that any other trait terms like "health span" or "aging" did not return any result. Interestingly, one of the most significant SNPs for PC2, rs13217066, has been linked to the *TBX18* gene involved in cardiac function, including regulation of diastolic blood pressure) [42], which is impacted with age [43].

Taken together the association results gained from the PC2 in GIFT highlighted multiple genes connected with various aging processes, supporting the speculation that this component maps on to the age at death variable. Interestingly, while GWAS have shown that *EHBP1* is associated with corpus callosum volume architecture [44], and delayed recall memory performance in AD patients [45], *EHBP1* is also a key regulator of lipophagy, i.e, the depletion of cytosolic lipid droplets, which reverses age-dependent loss of endoplasmic

reticulum (ER) proteostasis and enables ER morphological remodelling. This process, in turn, has been shown to drive beneficial metabolic changes and extends the lifespan in C. elegans [46].

Additionally, several genes highlighted affect mechanisms that have also been linked to dementia. *CLCN7* and *ERICHB* are associated with bone density [47] and maculopathy [48], processes linked to aging in the bone and eye respectively, and both low bone density and macular dystrophy have been linked to dementia [49, 50]. Similarly, genes *EPCIP* and *SGPL1* although have not been linked to age-related diseases, play key roles in processes known to influence aging. The *EPCIP* gene encodes the Exosomal Polycystin1 interacting protein, with is involved it the removal of senescent mitochondria, an organelle implicated in both aging [51, 52] and dementia [53], along with exosomes themselves [54]. Sphingolipid metabolism has also been linked with the aging process [55]; *SGPL1* encodes the enzyme, Sphingosine-1-phosphate lyase 1, which breaks down S1P. A build-up of S1P has been shown to be neurotoxic in mice [47], and associated with cognitive decline [56].

Interestingly other genes highlighted in PC2 have been directly linked dementia and frailty traits. *MEF2C* was found to be significantly associated with AD in the first GWAS [22]; however, this association has been missed in subsequent larger GWAS. Despite this *MEF2C* has been associated with aging, with evidence that its expression decreases with age and with it, its ability to “switch off” microglial to lower inflammation [57]. Further to this in a recent GWAS for frailty traits, *MEF2C* was again identified as being associated with poorer cognition [58]. In the same GWAS of frailty traits *CDH2* was highlighted as associated as a general factor [58]. Neurotrimin encoded by the *NTM* gene is known to play a role in neuronal cell adhesion and has an intronic SNP (rs73035809) associated with tau pathology burden [59].

CONCLUSION

Applying GIFT framework to human Alzheimer’s disease–related traits, its performance was compared with conventional GWAS using a modestly sized cohort (n = 563). By leveraging principal component–derived traits, we were able to disentangle AD pathology–related variation (PC1) from variation associated with age at death (PC2), thereby enabling a more nuanced exploration of genotype–phenotype relationships. Our findings demonstrate three key points. First, GIFT successfully recapitulated the canonical *APOE* locus associations identified by GWAS, including rs429358 and neighbouring variants within the *NECTIN2–TOMM40–APOE–APOC1* region. This concordance confirms that GIFT retains sensitivity to established, genome-wide significant signals and does not generate spurious associations. Second, GIFT identified additional SNPs beyond those detected by GWAS, including variants in genes with documented links to AD neuropathology. Third, GIFT revealed associations with age at death–related variation that were missed by GWAS. Although methodologically distinct, GIFT and GWAS are rooted in the same scientific tradition. Unlike many machine-learning approaches often characterised as “black boxes” due to limited interpretability, both GIFT and GWAS are transparent analytical frameworks in which assumptions and inferential procedures are explicitly defined. In this sense, GIFT constitutes a complementary and interpretable methodology capable of enhancing inference in contexts where large sample sizes are impractical but phenotypic resolution is high, thereby helping to mitigate longstanding power limitations in quantitative genomics.

Data Availability: Source data for this study are publicly available from Brains for Dementia Research (https://www.brainsfordementiaresearch.org/).

Acknowledgements: This work was funded by the Q-Life Project (Grant: PR1958).

Contributions: Conceptualization: CR; Methodology (theoretical formalisation of GIFT): CR, JAW; Software (coding and implementation): AP, SH, PG; Data Curation (BDR dataset): KJB; Clinical Investigation / Clinical Resources: BS, KK; Formal Analysis (results analysis): CR, BS, KK, KJB; Writing – Original Draft: CR, KJB; Writing – Review & Editing: BS, KK, KJB

Disclosures: The authors declare no competing interests.

REFERENCES

1. Bian, L. *et al.* Closing the gap between phenotyping and genotyping: review of advanced, image-based phenotyping technologies in forestry. *Annals of Forest Science* **79**, 22 (2022).
2. Kyratzi, P. *et al.* Investigative power of Genomic Informational Field Theory (GIFT) relative to GWAS for genotype-phenotype mapping. *Physiological Genomics* **56**, 791–806 (2024).
3. Rauch, C., Kyratzi, P., Blott, S., Bray, S. & Wattis, J. GIFT: new method for the genetic analysis of small gene effects involving small sample sizes. *Phys Biol* **20**, (2022).
4. Gadsby, S., Rauch, C. & Wattis, J. A. D. Identification of significant SNPs and the quantification of correlation using genomic informational field theory (GIFT). *Math Biosci* **393**, 109606 (2026).
5. Wattis, J. A. D., Bray, S. M., Kyratzi, P. & Rauch, C. Analysis of phenotype-genotype associations using genomic informational field theory (GIFT). *J Theor Biol* **548**, 111198 (2022).
6. Young, J. *et al.* Genome-wide association findings from the brains for dementia research cohort. *Neurobiology of Aging* **107**, 159–167 (2021).
7. Young, J. *et al.* Genome-wide association findings from the brains for dementia research cohort. *Neurobiology of Aging* **107**, 159–167 (2021).
8. Hyman, B. T. *et al.* National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement* **8**, 1–13 (2012).
9. Casper, J. *et al.* The UCSC Genome Browser database: 2026 update. *Nucleic Acids Res* **54**, D1331–D1335 (2026).
10. Phan, L. *et al.* The evolution of dbSNP: 25 years of impact in genomic research. *Nucleic Acids Res* **53**, D925–D931 (2025).
11. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742-015-0047-8 (2015).
12. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

13. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
14. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821–824 (2012).
15. Csardi, G. & Nepusz, T. The Igraph Software Package for Complex Network Research. *InterJournal* **Complex Systems**, 1695 (2005).
16. Grüning, B. *et al.* Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat Methods* **15**, 475–476 (2018).
17. Anaconda Software Distribution. Anaconda Inc. (2016).
18. Morris, J. C. Is Alzheimer's disease inevitable with age?: Lessons from clinicopathologic studies of healthy aging and very mild alzheimer's disease. *J Clin Invest* **104**, 1171–1173 (1999).
19. Brück, C. C. *et al.* Time to nursing home admission and death in people with dementia: systematic review and meta-analysis. *BMJ* **388**, e080636 (2025).
20. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research* **51**, D977–D985 (2023).
21. Shade, L. M. P. *et al.* GWAS of multiple neuropathology endophenotypes identifies new risk loci and provides insights into the genetic risk of dementia. *Nature Genetics* **56**, 2407–2421 (2024).
22. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45**, 1452–1458 (2013).
23. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics* **54**, 412–436 (2022).
24. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics* **51**, 404–413 (2019).
25. Wang, Q. *et al.* A Novel Missense Variant in SORBS2 Is Causative With Familial Alzheimer's Disease. *CNS Neurosci Ther* **31**, e70256 (2025).

26. Kang, M. *et al.* Genome-wide pleiotropy analysis of longitudinal blood pressure and harmonized cognitive performance measures. *Alzheimers Dement* 21, e70681 (2025).
27. Lee, J. H. *et al.* Genetic Modifiers of Age at Onset in Carriers of the G206A Mutation in PSEN1 With Familial Alzheimer Disease Among Caribbean Hispanics. *JAMA Neurol* 72, 1043–1051 (2015).
28. Mosquera-Heredia, M. I. *et al.* Long Non-Coding RNAs and Alzheimer's Disease: Towards Personalized Diagnosis. *Int J Mol Sci* 25, (2024).
29. Wang, C. *et al.* SOX7 interferes with β-catenin activity to promote neuronal apoptosis. *Eur J Neurosci* 41, 1430–1437 (2015).
30. Tosto, G. *et al.* Association of Variants in PINX1 and TREM2 With Late-Onset Alzheimer Disease. *JAMA Neurol* 76, 942–948 (2019).
31. Bottaro, D. P. *et al.* Identification of the Hepatocyte Growth Factor Receptor As the c-*met* Proto-Oncogene Product. *Science* 251, 802–804 (1991).
32. Hamasaki, H. *et al.* Down-regulation of MET in hippocampal neurons of Alzheimer's disease brains. *Neuropathology* 34, 284–290 (2014).
33. Matsumoto, K., Funakoshi, H., Takahashi, H. & Sakai, K. HGF-Met Pathway in Regeneration and Drug Discovery. *Biomedicines* 2, 275–300 (2014).
34. Liu, H., Luo, K. & Luo, D. Guanosine monophosphate reductase 1 is a potential therapeutic target for Alzheimer's disease. *Sci Rep* 8, 2759 (2018).
35. Qi, C. *et al.* Alzheimer's disease alters the transcriptomic profile of natural killer cells at single-cell resolution. *Front Immunol* 13, 1004885 (2022).
36. Andreyeva, A. *et al.* C-terminal fragment of N-cadherin accelerates synapse destabilization by amyloid-β. *Brain* 135, 2140–2154 (2012).
37. Ando, K. *et al.* N-cadherin regulates p38 MAPK signaling via association with JNK-associated leucine zipper protein: implications for neurodegeneration in Alzheimer disease. *J Biol Chem* 286, 7619–7628 (2011).

38. Lian, P. *et al.* Identification of metabolism-related subtypes and feature genes in Alzheimer's disease. *J Transl Med* **21**, 628 (2023).
39. Demange, P. A. *et al.* Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nat Genet* **53**, 35–44 (2021).
40. de la Fuente, J., Davies, G., Grotzinger, A. D., Tucker-Drob, E. M. & Deary, I. J. A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nat Hum Behav* **5**, 49–58 (2021).
41. Wright, K. M. *et al.* A Prospective Analysis of Genetic Variants Associated with Human Lifespan. *G3 (Bethesda)* **9**, 2863–2878 (2019).
42. Keaton, J. M. *et al.* Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat Genet* **56**, 778–791 (2024).
43. Safar, M. E. *et al.* Mean and yearly changes in blood pressure with age in the metabolic syndrome: the DESIR study. *Hypertension Research* **34**, 91–97 (2011).
44. Campbell, M. L. *et al.* Distributed genetic effects of the corpus callosum subregions suggest links to neuropsychiatric disorders and related traits. *Acta Neuropsychiatr* **37**, e23 (2023).
45. Homann, J. *et al.* Genome-Wide Association Study of Alzheimer's Disease Brain Imaging Biomarkers and Neuropsychological Phenotypes in the European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery Dataset. *Front Aging Neurosci* **14**, 840651 (2022).
46. Daniele, J. R. *et al.* UPR(ER) promotes lipophagy independent of chaperones to extend life span. *Sci Adv* **6**, eaaz1441 (2020).
47. Alam, S., Piazzesi, A., Abd El Fatah, M., Raucamp, M. & van Echten-Deckert, G. Neurodegeneration Caused by S1P-Lyase Deficiency Involves Calcium-Dependent Tau Pathology and Abnormal Histone Acetylation. *Cells* **9**, (2020).

48. Gurung, R. L. *et al.* Genome-Wide Association Study to Identify Genetic Variants Associated With Diabetic Maculopathy. *Invest Ophthalmol Vis Sci* **66**, 55 (2025).
49. Rozzini, L. *et al.* Cognitive dysfunction and age-related macular degeneration. *Am J Alzheimers Dis Other Demen* **29**, 256–262 (2014).
50. Xiao, T. *et al.* Association of Bone Mineral Density and Dementia: The Rotterdam Study. *Neurology* **100**, e2125–e2133 (2023).
51. Sun, N., Youle, R. J. & Finkel, T. The Mitochondrial Basis of Aging. *Mol Cell* **61**, 654–666 (2016).
52. Xu, X., Pang, Y. & Fan, X. Mitochondria in oxidative stress, inflammation and aging: from mechanisms to therapeutic advances. *Signal Transduction and Targeted Therapy* **10**, 190 (2025).
53. Wang, S. *et al.* Mitochondrial dysfunction in Alzheimer's disease: a key frontier for future targeted therapies. *Front Immunol* **15**, 1484373 (2024).
54. Safaei, S., Sohrabi, S., Zahmatkesh, P., Soltani-Zangbar, M. S. & Maleki, L. A. Exosomes in aging and age-related disorders: mechanisms, therapeutic potentials, and challenges. *J Transl Med* **23**, 1423 (2025).
55. Li, S. & Kim, H.-E. Implications of Sphingolipids on Aging and Age-Related Diseases. *Front Aging* **2**, 797320 (2021).
56. Moseholm, K. F. *et al.* Circulating sphingolipids in relation to cognitive decline and incident dementia: The Cardiovascular Health Study. *Alzheimers Dement (Amst)* **16**, e12623 (2024).
57. Deczkowska, A. *et al.* Mef2C restrains microglial inflammatory response and is lost in brain ageing in an IFN-I-dependent manner. *Nature Communications* **8**, 717 (2017).
58. Foote, I. F. *et al.* Uncovering the multivariate genetic architecture of frailty with genomic structural equation modeling. *Nature Genetics* **57**, 1848–1859 (2025).

59. White, C. C. *et al.* A genome-wide investigation of clinicopathologic endophenotypes uncovers a new susceptibility locus for tau pathology at Neurotrimin (NTM). *Alzheimer's & Dementia* **17**, e051682 (2021).

60. Loika, Y., Loiko, E., Culminskaya, I. & Kulminski, A. M. Exome-Wide Association Study Identified Clusters of Pleiotropic Genetic Associations with Alzheimer's Disease and Thirteen Cardiovascular Traits. *Genes (Basel)* **14**, (2023).

61. Abraham, R. *et al.* A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics* **1**, 44 (2008).

62. Wang, H. *et al.* Similar Genetic Architecture of Alzheimer's Disease and Differential APOE Effect Between Sexes. *Front Aging Neurosci* **13**, 674318 (2021).

63. Koskeridis, F. *et al.* Multi-trait association analysis reveals shared genetic loci between Alzheimer's disease and cardiovascular traits. *Nat Commun* **15**, 9827 (2024).

64. Nazarian, A., Yashin, A. I. & Kulminski, A. M. Genome-wide analysis of genetic predisposition to Alzheimer's disease and related sex disparities. *Alzheimers Res Ther* **11**, 5 (2019).

65. Marioni, R. E. *et al.* GWAS on family history of Alzheimer's disease. *Transl Psychiatry* **8**, 99 (2018).

66. Lo, M.-T. *et al.* Identification of genetic heterogeneity of Alzheimer's disease across age. *Neurobiol Aging* **84**, 243.e1-243.e9 (2019).

67. Willett, J. D. S. *et al.* Identification of 16 novel Alzheimer's disease loci using multi-ancestry meta-analyses. *Alzheimers Dement* **21**, e14592 (2025).

68. Hollingworth, P. *et al.* Genome-wide association study of Alzheimer's disease with psychotic symptoms. *Mol Psychiatry* **17**, 1316–1327 (2012).

69. Ge, Y.-J. *et al.* Genome-wide meta-analysis identifies ancestry-specific loci for Alzheimer's disease. *Alzheimers Dement* **20**, 6243–6256 (2024).

70. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1276–1282 (2021).
71. Hong, S. *et al.* Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD Multimodal Biomarker Discovery dataset. *Transl Psychiatry* **10**, 403 (2020).
72. Wang, P. *et al.* Genome-wide association studies identify novel loci in rapidly progressive Alzheimer's disease. *Alzheimers Dement* **20**, 2034–2046 (2024).
73. Moreno-Grau, S. *et al.* Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: The GR@ACE project. *Alzheimers Dement* **15**, 1333–1347 (2019).
74. Nazarian, A., Arbeev, K. G., Yashkin, A. P. & Kulminski, A. M. Genetic heterogeneity of Alzheimer's disease in subjects with and without hypertension. *Geroscience* **41**, 137–154 (2019).
75. Talyansky, S., Le Guen, Y., Kasireddy, N., Belloy, M. E. & Greicius, M. D. APOE-ε4 and BIN1 increase risk of Alzheimer's disease pathology but not specifically of Lewy body pathology. *Acta Neuropathol Commun* **11**, 149 (2023).
76. Akgun, B. *et al.* Genome-wide association analysis and admixture mapping in a Puerto Rican cohort supports an Alzheimer disease risk locus on chromosome 12. *Front Aging Neurosci* **16**, 1459796 (2024).
77. Rajabli, F. *et al.* Multi-ancestry genome-wide meta-analysis of 56,241 individuals identifies known and novel cross-population and ancestry-specific associations as novel risk loci for Alzheimer's disease. *Genome Biol* **26**, 210 (2025).
78. Harper, J. D. *et al.* Genome-Wide Association Study of Incident Dementia in a Community-Based Sample of Older Subjects. *J Alzheimers Dis* **88**, 787–798 (2022).
79. Dacey, R. *et al.* Alzheimer's disease multi-ancestry genome-wide interaction and stratified study with smoking. *Alzheimers Dement* **21**, e70922 (2025).

80. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).

81. Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet* **53**, 392–402 (2021).

82. He, L. *et al.* Exome-wide age-of-onset analysis reveals exonic variants in ERN1 and SPPL2C associated with Alzheimer's disease. *Transl Psychiatry* **11**, 146 (2021).

83. Zhang, Y.-R. *et al.* Whole exome sequencing analyses identified novel genes for Alzheimer's disease and related dementia. *Alzheimers Dement* **20**, 7062–7078 (2024).

84. Sherva, R. *et al.* African ancestry GWAS of dementia in a large military cohort identifies significant risk loci. *Mol Psychiatry* **28**, 1293–1302 (2023).

Table 1: Information on significant SNPs determined by GWAS

								Phenotype significance (-Log(P_{GWAS}))			
CHR	**POS**	**rsID**	**REF**	**ALT**	**MAF**	**Gene name**	**Biotype**	**$PC1_{AD}$**	**$PC1_{AD+Age}$**	**$PC2_{AD+Age}$**	**GWAS catalog**
7	24171927	rs6952104	C	G	0.467227	NA	NA	0.875	1.130	**5.264**	
19	45382034	rs6859*	A	G	0.454698	NECTIN2	Pc	**6.229**	**6.400**	0.499	[60–66]
	45388130	rs34342646*	G	A	0.226589	NECTIN2, LNCOB1	Pc, lncRNA	**16.372**	**17.080**	1.820	[67]
	45394336	rs71352238*	T	G	0.226968	NA	NA	**16.532**	**17.262**	1.928	[64,67]
	45395909	rs34404554*	C	G	0.222408	TOMM40	Pc	**17.085**	**17.842**	1.960	[67]
	45396219	rs157582*	C	A	0.29916	TOMM40	Pc	**19.076**	**19.496**	0.598	[64,67,68]
	45401666	rs8106922*	A	G	0.355351	TOMM40	Pc	**6.606**	**6.483**	0.459	[64,67]
	45403412	rs1160985*	C	A	0.380435	TOMM40	Pc	**8.172**	**8.135**	0.129	[64,67]
	45410002	rs769449*	G	A	0.213211	APOE	Pc	**24.663**	**25.653**	2.012	[64,67,69]
	45411941	rs429358*	T	C	0.312919	APOE	Pc	**20.144**	**20.979**	1.667	[62,64,66,67,69–84]
	45421254	rs12721046*	G	A	0.237815	APOC1	Pc	**21.787**	**22.598**	1.681	[67]

Numerical values are shown in bold where they exceed the 5% Bonferroni-corrected significance thresholds ($PC1_{AD}$: 5.108; $PC1_{AD+Age}$: 5.111; $PC2_{AD+Age}$: 5.116). An asterisk (*) next to the rsID indicates variants that reached significance in both GWAS and GIFT analyses.

Table 2: Information on significant SNPs determined by GIFT

CHR	POS	rsID	REF	ALT	MAF	Gene name	Biotype	Phenotype significance (-Log(P_{GIFT}))		
								$PC1_{AD}$	$PC1_{AD+Age}$	$PC2_{AD+Age}$
2	51317471	rs6734633	G	A	0.301003	NRXN1-DT, NRXN1	lncRNA, Pc	5.344	**6.731**	2.053
	59903052	rs2110330	A	G	0.420435	NA	NA	1.206	1.592	**6.329**
	63007059	rs10153716	A	G	0.305695	EHBP1	Pc	0.466	0.971	**5.929**
	63073194	rs11682530	G	A	0.307886	EHBP1	Pc	0.511	1.146	**5.993**
	123524253	rs6541859	A	G	0.156355	NA	NA	4.294	**5.912**	1.536
	146737198	rs16825854	G	A	0.116415	NA	NA	0.429	0.960	**6.487**
	146835932	rs10496995	C	A	0.119765	NA	NA	0.705	1.208	**6.371**
	168710172	rs10200961	A	G	0.124161	B3GALT1, B3GALT1-AS1	Pc, lncRNA	**5.947**	**6.308**	3.864
	205177698	rs7603542	C	A	0.11204	NA	NA	**5.922**	**6.525**	1.103
4	16281681	rs11721706	A	G	0.144295	TAPT1-AS1	lncRNA	3.601	**6.136**	1.070
	16365531	rs16893404	G	A	0.142977	NA	NA	3.405	**5.960**	0.983
	26501114	rs11733591	C	A	0.069398	NA	NA	**6.577**	**6.706**	1.836
	186714443	rs10866283	G	A	0.098662	SORBS2	Pc	5.299	**6.533**	1.239
5	88009593	rs254778	G	A	0.313758	MEF2C-AS2	lncRNA	0.607	1.079	**6.2523**
6	85347257	rs13217066	C	A	0.406355	NA	NA	2.335	4.057	**7.4584**
7	103659343	rs17158237	C	A	0.109532	NA	NA	5.760	**6.6287**	0.862
	103670208	rs7799089	A	G	0.11204	NA	NA	5.155	**5.9452**	0.892
	116435123	rs13239139	A	G	0.175042	MET	Pc	4.446	**6.846**	1.361
	136782819	rs11761344	C	A	0.123746	NA	NA	5.609	**6.918**	1.481
8	10627844	rs7826180	G	A	0.239531	SOX7-AS1, PINX1	lncRNA, Pc	**6.372**	5.695	1.396
	25557962	rs11997651	T	G	0.275084	NA	NA	1.685	1.917	**6.282**
	26867904	rs2874755	C	A	0.0752508	NA	NA	0.960	1.480	**6.332**
	27057879	rs12548107	G	A	0.0643813	NA	NA	0.821	0.977	**6.769**
	53445338	rs13282005	T	C	0.10134	NA	NA	1.686	1.866	**6.390**
10	3815292	rs2171302	C	A	0.16806	NA	NA	0.569	0.819	**6.233**
	72713252	rs827284	T	A	0.0991597	SGPL1, LINC02622	Pc, lncRNA	3.605	4.331	**6.008**
	115788094	rs11196597	G	A	0.107023	NA	NA	4.629	**5.947**	0.781
11	131355411	rs7951332	G	A	0.0543478	NTM	Pc	1.000	0.962	**6.230**
12	4475194	rs10744643	C	G	0.383585	NA	NA	0.732	0.994	**7.201**
	10462516	rs2270238	C	A	0.328859	KLRD1	Pc	4.448	**5.918**	1.292
	63129713	rs7312352	A	G	0.119565	PPM1H	Pc	5.676	**6.023**	1.747
	101375823	rs17030838	A	G	0.184783	ANO4	Pc	3.353	**6.304**	1.727
13	46162672	rs7329523	T	G	0.132107	ERICH6B	Pc	2.127	2.230	**6.661**
16	1498951	rs12932780	C	A	0.101171	CLCN7	Pc	2.703	3.208	**7.454**
	86347358	rs2696819	G	A	0.262982	NA	NA	1.328	1.694	**6.766**

	86357163	rs8059183	G	A	0.218487	NA	NA	0.946	1.138	**5.952**
17	50617026	rs1492230	A	G	0.330268	NA	NA	1.487	2.378	**6.167**
18	25569058	rs2298574	A	G	0.103015	CDH2	Pc	3.787	**6.027**	1.321
	41183021	rs1830221	G	A	0.352349	NA	NA	1.481	1.852	**6.049**
19	34595690	rs11881397	G	A	0.0854271	NA	NA	2.615	3.235	**6.928**
	45382034	rs6859*	A	G	0.454698	NECTIN2	Pc	**5.996**	**6.947**	2.644
	45388130	rs34342646*	G	A	0.226589	NECTIN2, LNCOB1	Pc, lncRNA	**12.085**	**13.494**	**6.473**
	45394336	rs71352238*	T	G	0.226968	NA	NA	**12.191**	**13.387**	**6.411**
	45395909	rs34404554*	C	G	0.222408	TOMM40	Pc	**12.828**	**14.233**	**6.411**
	45396219	rs157582*	C	A	0.29916	TOMM40	Pc	**13.587**	**15.015**	4.817
	45401666	rs8106922*	A	G	0.355351	TOMM40	Pc	5.750	**6.192**	2.895
	45403412	rs1160985*	C	A	0.380435	TOMM40	Pc	**7.347**	**8.108**	2.723
	45410002	rs769449*	G	A	0.213211	APOE	Pc	**19.370**	**21.597**	**6.823**
	45411941	rs429358*	T	C	0.312919	APOE	Pc	**17.512**	**19.446**	4.974
	45421254	rs12721046*	G	A	0.237815	APOC1	Pc	**16.565**	**18.396**	5.861
20	17775259	rs6080834	G	A	0.257525	NA	NA	2.173	2.215	**7.526**
21	24951271	rs2828286	A	G	0.144054	NA	NA	4.656	**6.547**	0.675
	34279458	rs7283473	G	A	0.31689	EPCIP-AS1	lncRNA	1.532	2.089	**6.348**
	34292163	rs7277038	C	A	0.317726	EPCIP-AS1	lncRNA	1.326	1.837	**6.739**
22	43049014	rs5758837	G	A	0.371237	CYB5R3	Pc	4.329	**6.226**	1.369

Numerical values are shown in bold where they exceed the 5% Bonferroni-corrected significance thresholds ($PC1_{AD}$: 5.762; $PC1_{AD+Age}$: 5.909 and $PC2_{AD+Age}$: 5.912). An asterisk (*) next to the rsID indicates variants that reached significance in both GWAS and GIFT analyses.

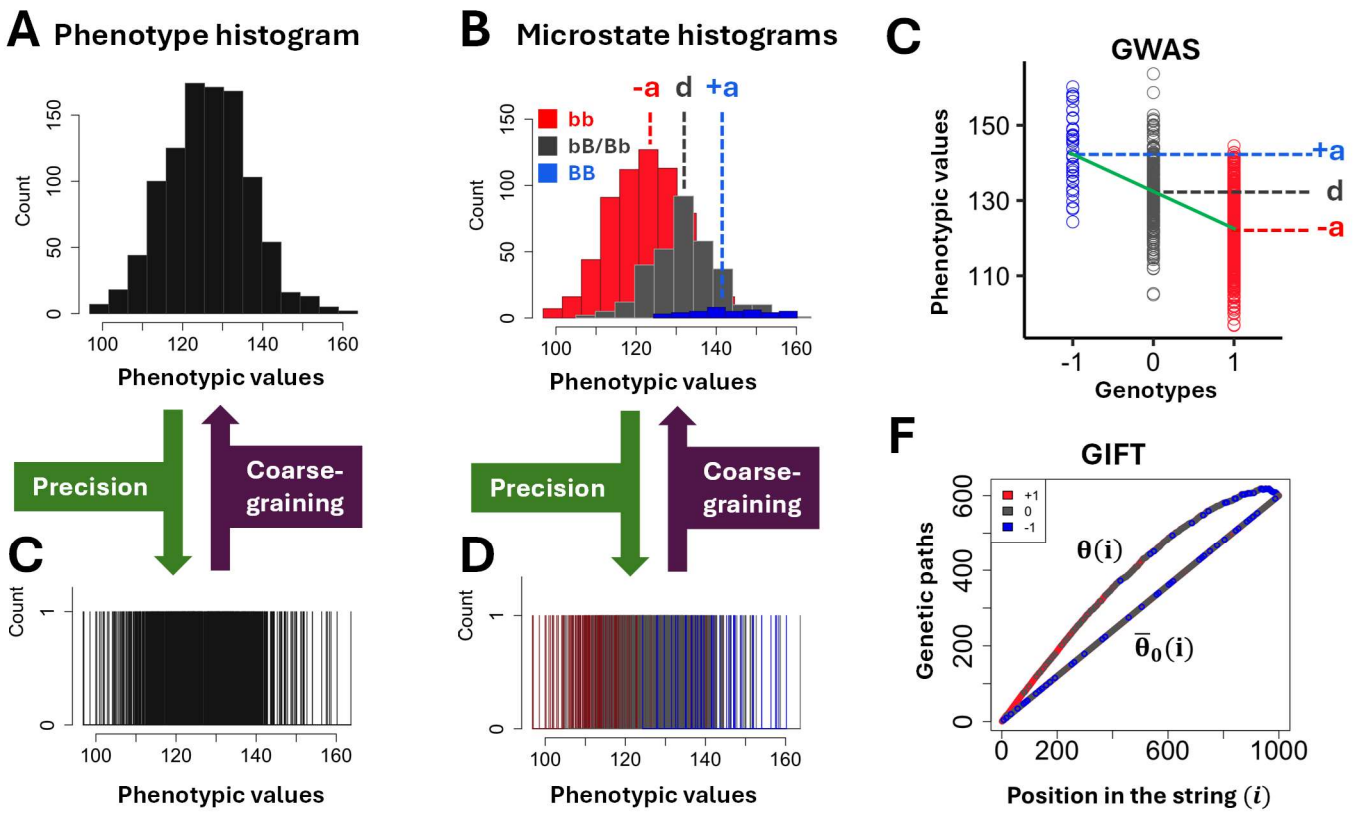

Figure 1: GWAS and GIFT simulation using 1,000 individuals, with a gene effect (2a) equal to the standard deviation of the phenotype distribution, zero dominance, and genotype frequencies consistent with Hardy–Weinberg equilibrium (HWE). **(A)** Representation of the phenotype distribution using a histogram (count plot). **(B)** Representation of genotype frequencies using count plots. In biallelic (b, B) organisms, three genotypes are observed: bb, bB/Bb, and BB. **(C)** In the GWAS framework, genotype–phenotype associations are evaluated by comparing the mean phenotypic values associated with each genotype (−a, d, +a), allowing estimation of the statistical significance of the association, the gene effect size (2a), and dominance (d). **(D–E)** As measurement precision increases, fewer data points fall within each bin, and at full precision the histogram effectively resolves into barcode-like representations. Crucially, coarse-graining GIFT' s detailed outputs (barcodes) recover GWAS representational framework **(F)** GIFT determines genotype–phenotype associations by evaluating the probability of specific genotype configurations, introducing the concept of genetic paths to infer statistical significance.

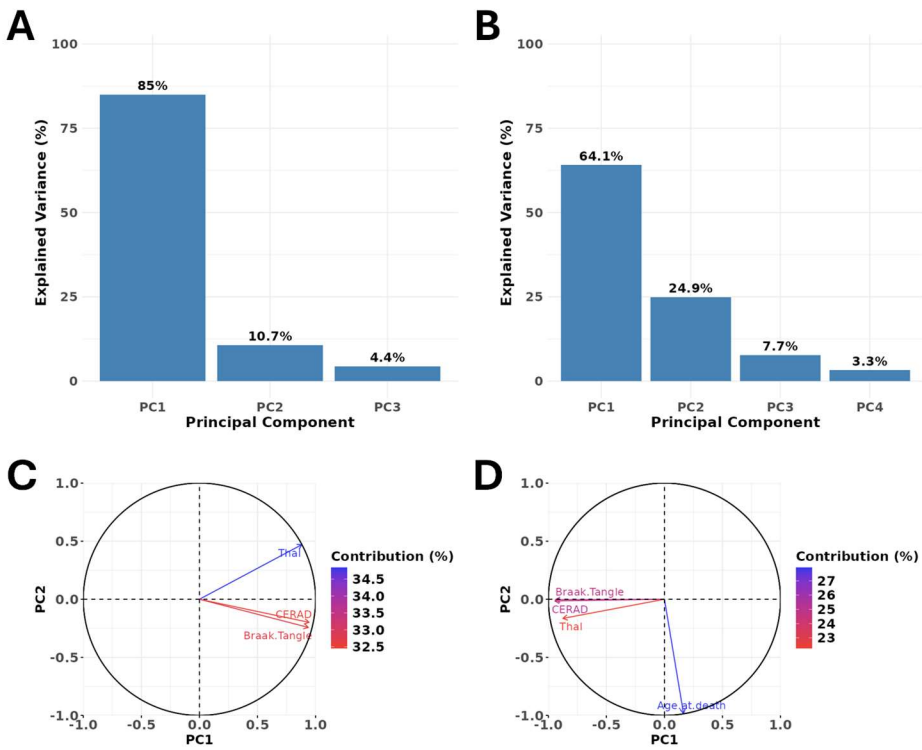

Figure 2: Principal components analyses. (A) Fraction of variances explained by the different components using the AD neuropathology variables dataset. **(B)** Fraction of variances explained by the different components using the AD neuropathology and age at death variables. **(C)** Representation of AD neuropathology vectors along the principal components PC1 and PC2. **(D)** Representation of AD neuropathology and age at death vectors along the principal components PC1 and PC2. Note: coloured labels on each bar in (C) and (D) show rounded calculations of variance explained by the components.

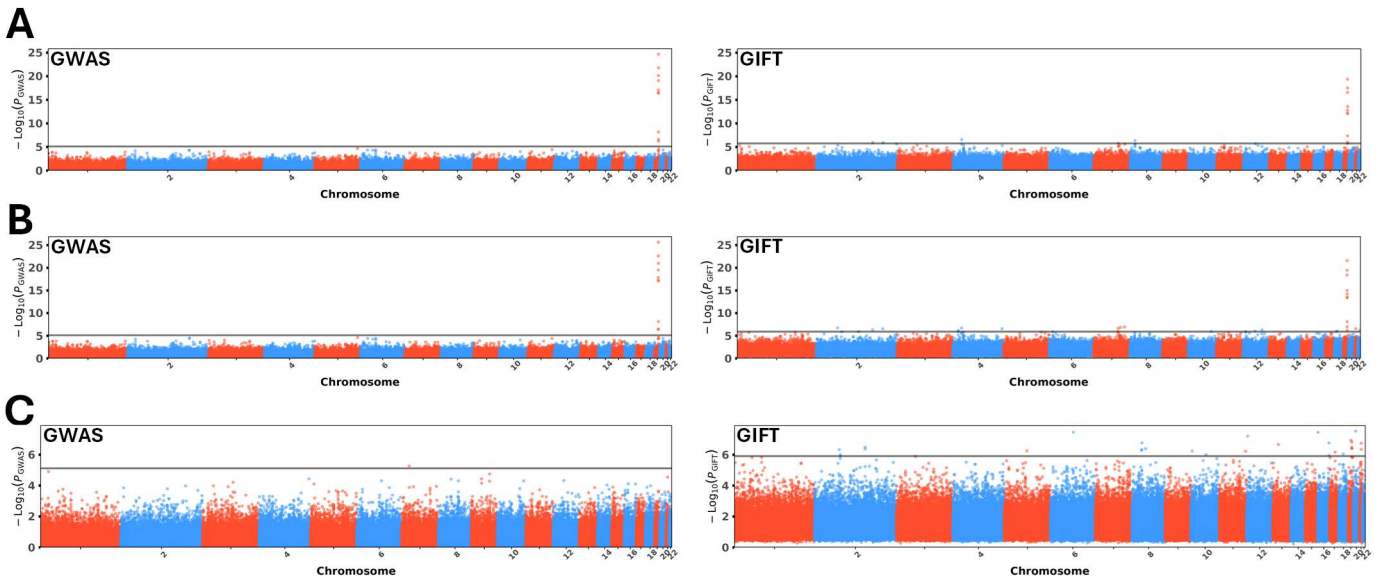

Figure 3: GWAS and GIFT Manhattan plots: (A) Genome-wide association for AD pathology variables (PC1). **(B)** Genome-wide association for AD pathology and age at death variables (PC1). **(C)** Genome-wide association for AD pathology and age at death variables (PC2). The horizontal grey line corresponds to Bonferroni correction (5%).

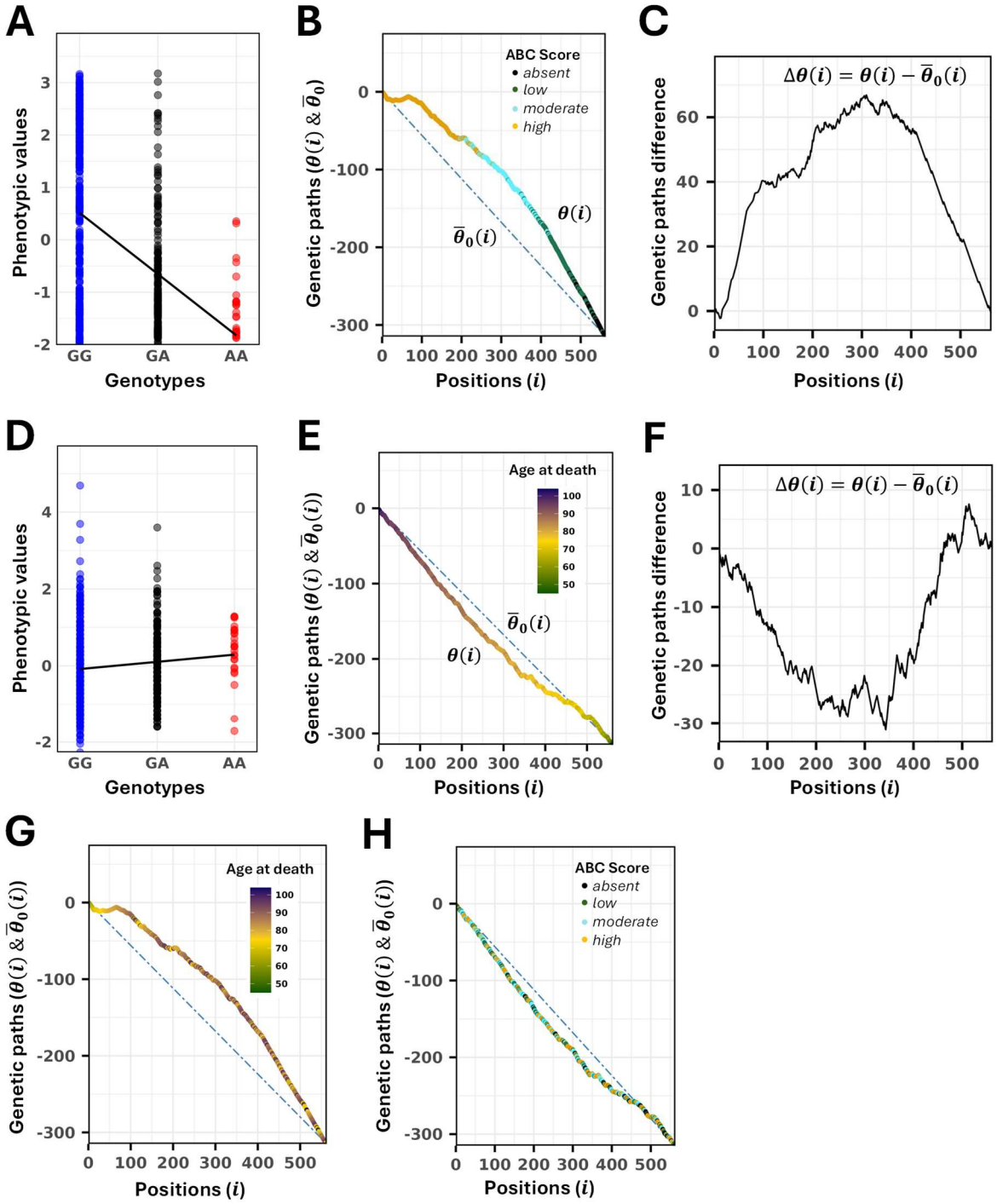

Figure 4: (A) GWAS representation of the allelic substitution in rs769449 (Chr19) based on $PC1_{AD+Age}$ from the AD pathology variables and age at death dataset. **(B)** GIFT representation of rs769449 using on that $PC1_{AD+Age}$ from the AD pathology variables and age at death dataset. Two genetic paths are plotted, one associated with the phenotypic values ($\theta(i)$) and the second not associated with the phenotypic values ($\bar{\theta}_0(i)$). **(C)** GIFT representation of the difference between the genetic paths determined in (B) used to compute the level of significance underscoring the association. **(D)** GWAS representation of the allelic substitution of rs769449 (Chr19) based on that $PC2_{AD+Age}$ from the AD pathology variables and age at death dataset. **(E)** GIFT representation of rs769449 using on that $PC2_{AD+Age}$ from the AD pathology variables and age at death dataset. Two genetic paths are plotted, one associated with the phenotypic values ($\theta(i)$) and the second not associated with the phenotypic values ($\theta_0(i)$). **(F)** GIFT representation of the difference between the genetic paths determined in (E) used to compute the level of significance underscoring the association. **(G)** Genetic paths from (B) overlaid with age at death measures. **(H)** Genetic paths from (E) overlaid with AD pathology variables.