Multivariate genotypic analyses that identify specific genotypes to characterize disease and control groups in ADNI

Derek Beaton*

Rotman Research Institute

Baycrest Health Sciences

3560 Bathurst Street

Toronto, ON, Canada M6A 2E1

Jenny R. Rieck

Rotman Research Institute

Baycrest Health Sciences

3560 Bathurst Street

Toronto, ON, Canada M6A 2E1

Fahd Alhazmi

School of Behavioral and Brain Sciences

The University of Texas at Dallas

MS: GR4.1

800 West Campbell Rd.

Richardson, TX 75080 USA

ADNI**

Alzheimer's Disease Neuroimaging Initiative

Hervé Abdi*

School of Behavioral and Brain Sciences

The University of Texas at Dallas MS: GR4.1,

800 West Campbell Rd.

Richardson, TX 75080 USA

Author Note

The majority of the work in this article began when all authors were at The University of Texas at Dallas. DB and JRR are currently at the Rotman Research Institute and FA is at CUNY.

*Corresponding authors

**Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://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 list of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Acknowledgements

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.

Abstract

INTRODUCTION: Genetic contributions to Alzheimer's Disease (AD) are likely polygenic and not necessarily explained by the additive model alone. In order to better understand the genetics of AD we require statistical techniques to address both polygenic and possible non-additive effects.

METHODS: We used partial least squares-correspondence analysis (PLS-CA). PLS-CA was designed to detect multivariate genotypic effects. We used ADNI-1 (N = 756) as a discovery sample with two forms of PLS-CA: diagnosis-based and *APOE*-based. We used ADNI-2 (N = 791) as a validation sample with a diagnosis-based PLS-CA.

RESULTS: With PLS-CA we identified some expected genotypic effects (e.g., *APOE*/*TOMM40*, and *APP*) and a number of novel effects. Novel effects include for examples: risk-associated genotypes in *RBFOX1* and *GPC6* and control-associated (possibly protective) genotypes in *PTPN14* and *CPNE5*.

DISCUSSION: To better understand the genetic contributions to AD we require techniques designed to detect complex (multivariate, genotypic) effects.

Keywords: Genome-wide pattern analysis, partial least squares, correspondence analysis, genotypic model, discriminant analysis, statistical methods

Introduction

The discovery of the functional roles of the *APOE* genotypes in Alzheimer's disease (AD) [1–3] initiated the exploration of the genetic contributions to late-onset or sporadic AD and its many associated traits. Since then numerous studies have identified many genes linked to AD such as *SORL1* [4], *EPHA1* [5], *CLU* and *PICALM* [6], all of which have shown effects at much larger scales [7]. More recently studies have focused on non-Caucasian [8] or trans-ethnic [9] populations in order to reveal genetic contributions to AD across ethnicities. Even though the genetic contributions to AD are becoming well understood—as documented by several comprehensive reviews [10–12]—there are still some controversial findings. For example, the work reported in [13] showed effects of rare variants in *PLD3*, and some follow up studies showed both replication [14] and no replication [15] of *PLD3* effects. It is clear that the genetics of late-onset or sporadic AD are very complex and almost certainly polygenic [16].

Recent work in late-onset or sporadic AD has emphasized polygenic [17,18], epistatic [19], and non-linear or non-additive [20] effects. These studies suggest that simple recessive, dominant, or additive models do not adequately explain effects in AD; an effect that is well known in other complex traits and diseases [21]. But if the genetics of AD are complex why are we still using methods (i.e., univariate) with adherence to assumptions (i.e., additivity) unsuited to detect complex effects?

Here we present a method designed specifically to help address some of the genetic complexities (i.e., polygenicity, non-additivity) in AD: partial least squares-correspondence analysis (PLS-CA). PLS-CA is a multivariate technique that detects rather than assumes genetic effects (e.g., additive vs. recessive). First we provide a review in order to illustrate the complexity of genetics in AD, followed by an explanation of our method and finally outline the

remainder of our paper.

Complex genetic effects in AD

The AD literature documents a number of well-established complex relationships not explained by simple effects (i.e., additive, recessive, or dominant). For examples, *APOE* shows different risk factors for AD based on age [22] and *TOMM40* is co-dominant [23]. Additionally, *APOE* genotypes show different trajectories of CSF tau and beta-amyloid (Aβ) deposition [24], and non-linear effects have also been observed in tau patterns across different brain regions based on the *MAPT* gene [25]. Furthermore, the same genotype can express varying patterns dependent on the phenotype of interest, as illustrated, for example, by Figure 2 in [26] which shows opposite genotypic effects for functional vs. structural MRI analyses. Patterns become even more complex with multiple markers [27,28] and haplotypes [29]. Overall, these results suggest that different markers are likely to express different inheritance patterns depending on the trait, phenotype, or assessment used in the study.

Issues with the additive model

With the advent of genome-wide studies, single nucleotide polymorphisms (SNPs) have almost exclusively been analyzed with the additive model. The additive model transforms a SNP from base pair letters into a count based on, usually, the number of minor alleles: a major homozygote is “0,” a heterozygote is “1,” and a minor homozygote is “2.” This {0,1,2} coding scheme has been used in a wide array of studies such as the onset of AD [30], and cerebrospinal fluid (CSF) tau levels [31], even though we generally acknowledge that the standard approaches cannot detect non-additive or complex effects [32,33]. Furthermore, with open resources such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI), researchers have the opportunity to

analyze genome-wide data in a variety of ways. However, the ADNI genome-wide data have been almost exclusively analyzed with the additive model [34–39].

The additive model has become popular because it is viewed as both a practical approach [40] and a likely model for the inheritance of complex traits [41]. However, the strict assumptions of additivity and linearity do not always hold (see responses and comments to [41] in [42,43]). Furthermore, the additive model can produce ambiguity across samples or studies. This can be directly observed in the ADNI-1 vs. the ADNI-GO/2 data with thousands of SNPs. For one example: rs308076 in ADNI-1 'T' is the minor allele (MAF = 47.5%) vs. ADNI-GO/2 where 'C' is the minor allele (MAF = 43.3%); thus when the cohorts are kept separate from one another the '0' and '2' are not the same genotypes across both studies. Furthermore, this ambiguity of '0' and '2' could possibly lead to the misinterpretation, or the dismissal of effects especially between samples or replications. For examples in AD: in [44] the authors reported "direction changes" in their own replication analyses, whereas in [45] the authors report effects in the opposite direction in their attempt to replicate work in [13]. A possibly extreme case of this can be observed in substance use disorders in [46] where direction changes were interpreted as "false positives."

The *a priori* choice of a model that does not match the true inheritance pattern can be problematic (e.g., using an additive model when the real effect is recessive). For candidate studies the work in [47] showed the consequences of such *a priori* choices, and recommended two strategies: (1) test all models (e.g., additive, dominant, and recessive) and apply appropriate multiple test corrections, or (2) use a more general model (e.g., co-dominant, or genotypic). Both strategies provided comparable power (see Figs. 1 and 2 in [47]). Furthermore more recent work (in cholesterol) has shown that the genotypic ("full") model is better than the additive model to

detect genetic contributions to complex traits [48]. While either strategy described in [47] is feasible in candidate studies, to test multiple models (e.g., additive, dominant, and recessive) in GWAS would lead to even more conservative—and thus unrealistic—thresholds for tests.

Multivariate analysis with the genotypic model

In this study we investigate the ADNI genome-wide data (both ADNI-1 and ADNI-GO/2) in a more general way (i.e., analogous to a genotypic model; see *Methods*) and evaluated the SNPs and their relations without the assumptions (e.g., linearity, magnitude, and direction of effect, and independence of the SNPs) of any particular model. Therefore, in our study we not only focus on genotypes, but we also use a multivariate approach to assess which genotypes (1) may work together, and (2) are most associated with the each of the groups in our study (i.e., AD, Mild Cognitive Impairment [MCI], and control [CON]).

To achieve this, we used a multivariate technique designed specifically for categorical data called partial least squares-correspondence analysis (PLS-CA, [49]). PLS-CA is a generalization of two different families of techniques: partial least squares (PLS; as in neuroimaging: [50,51]), and correspondence analysis (CA; [52–55]). PLS-CA analyzes the relationships between two tables of categorical or mixed data. In our study, we used two specific forms of PLS-CA called discriminant PLS-CA (a.k.a., mean-centered PLS-CA or discriminant correspondence analysis; [56]) and "seed" PLS-CA (see *Methods* and [49] for more details). Furthermore, as a direct analog to the neuroimaging literature, our approach is akin to "genome-wide pattern analysis" that identifies how multiple genotypes across the genome (à la multiple voxels across the brain) may work together.

Our paper is outlined as follows. In *Methods*, we briefly describe the ADNI project, its participants, and the measures we used in this study. We then describe the preprocessing and

analytical strategies and tools. Finally, we detail the two "phases" of our study. Phase 1 ("discovery"; ADNI-1) includes two genome-wide association analyses: one where the association is with clinical diagnosis (mean-centered/discriminant) and the other where the association is with *APOE* genotype ("seed"). Phase 2 ("validation") used the genotypes identified in Phase 1 to test how well those genotypes (discovered in ADNI-1) could predict group relationship (AD, MCI, or CON) in a left-out sample (validated in ADNI-GO/2). Then, the *Results* section reviews the findings from each of the Phases. The *Discussion* section first provides a discussion of the findings within each phase, followed by a general discussion, limitations, and conclusions.

Methods

Data

Data used in the preparation of this article come from the ADNI database (http://adni.loni.usc.edu). The ADNI project was launched in 2003 as a public-private funding partnership and includes public funding by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and the Food and Drug Administration. The primary goal of ADNI has been to collect a wide variety of measures to assess the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. Michael W. Weiner, MD (VA Medical Center and University of California at San Francisco) is the ADNI Principal Investigator. Subjects have been recruited from over 50 sites across the U.S. and Canada. For up-to-date information, see www.adni-info.org.

The following studies included genomic data from ADNI-1 and ADNI-GO/2 (in addition to *APOE* genotype, and diagnostic data from the `ADNIMERGE` R package). Note that the *APOE*

genotype data were collected separately from genome-wide data and that the ADNI genome-wide data chips were different in ADNI-1 and in ADNI-GO/2. ADNI-1 used the Illumina Human610-Quad BeadChip while ADNI-GO/2 used the Illumina HumanOmniExpress BeadChip. For more details on ADNI genotyping, see http://adni.loni.usc.edu/data-samples/genetic-data/

Participants

We only included participants that had genome-wide data and obtained a final total of 756 participants from ADNI-1 (AD = 344, MCI = 204, CON = 208) and 791 participants from ADNI-GO/2 (AD = 203, MCI = 319, CON = 269). The MCI category was subdivided into two groups in ADNI-GO/2 (early and late MCI); we combined both into a single MCI "diagnostic" category to align with ADNI-1. Table 1 contains race, ethnicity, sex, and diagnostic distributions for ADNI-1 and ADNI-GO/2. We used the `ADNIMERGE` package to extract *APOE* genotype (for ADNI-1 only), diagnostic, and demographic information. In our analyses, we opted to use the full set of participants and correct for stratification effects as opposed to restricting our analyses to a particular ethnic and racial group in order to identify genotypes that contribute to AD across race and ethnicity.

Table 1

Demographics of ADNI 1 and 2.

(a) Race (non-Hispanic/Latino)

	ADNI1	ADNI-GO/2
White	703 (685)	732 (700)
Asian	12 (12)	12 (12)
Black	37 (36)	30 (30)
American Indian/Alaskan	1 (1)	2 (2)
Hawaiian/Other Pacific Islander	0 (0)	2 (2)
Unknown/More than 1	3 (1)	13 (7)
Total	***756*** (***735***)	***791*** (***753***)

(b) Sex

	ADNI1	ADNI-GO/2
Female	308	367
Male	448	424
Total	***756***	***791***

(c) Diagnosis

	ADNI1	ADNI-GO/2
Control	208	269
Mild Cognitive Impairment	204	319
Alzheimer's Disease	344	203
Total	***756***	***791***

NOTE. Demographics information for ADNI 1 and ADNI-GO/2. All demographics information retrieved from the ADNIMERGE package. Diagnosis (Dx) was the last available diagnosis. Table (a) shows the distribution of race with the number of non-Hispanic/Latino in parentheses. "Unknown" and "More than 1" were combined into a single category. Table (b) shows the distribution of sex for each phase of ADNI. Table (c) shows the distribution of latest-available diagnosis (at the time of analyses) for each phase of ADNI.

Statistical techniques

Most preprocessing, analyses, and graphics were performed primarily in `R` [57] and the `ExPosition`, `TExPosition`, and `TInPosition` packages [58,59] in `R`. Some in-house `MATLAB` (Mathworks Inc., Natick, MA) code was used for resampling.

Data for analyses—e.g., SNPs, diagnoses, *APOE* E4 status—were recoded into a disjunctive format described in Table 2. Because we treat our data categorically we required particular multivariate techniques designed specifically for categorical data. We used multiple correspondence analysis (MCA) and two forms of partial least squares-correspondence analysis (PLS-CA): discriminant PLS-CA and seed PLS-CA [49]. We briefly describe these techniques here but also provide more details where necessary.

Table 2

Nominal and disjunctive formats of data.

(a) Nominal

	SNP1 (with minor homozygote > 5%)	SNP2 (with minor homozygote < 5%)
Subj.1	AG	CA
Subj.2	AA	CA
…	…	…
Subj.i	AG	CC
...	…	…
Subj.I-1	<NA>	AA
Subj.I	GG	AA

(b) Disjunctive SNPs

	SNP1 (minor homozygote > 5%			SNP2 (minor homozygote < 5%)	
	AG	AA	GG	CA+CC	AA
Subj.1	1	0	0	1	0
Subj.2	0	1	0	1	0
…	…	…	…	…	…
Subj.i	1	0	0	0	0
…	…	…	…	…	…
Subj.I-1	.2	.7	.1	0	1
Subj.I	0	0	1	0	1

(c) Dx and *APOE*

	Dx	#*APOE* E4
Subj.1	AD	2
Subj.2	AD	1
…	…	…
Subj.i	MCI	1
...	…	…
Subj.I-1	MCI	2
Subj.I	CN	0

(d) Disjunctive Dx and *APOE*

	Dx			# *APOE* E4 Alleles		
	AD	MCI	CN	2	1	0
Subj.1	1	0	0	1	0	0
Subj.2	1	0	0	0	1	0
…	…	…	…	…	…	…
Subj.i	0	1	0	0	1	0
…	…	…	…	…	…	…
Subj.I-1	0	1	0	1	0	0
Subj.I	0	0	1	0	0	1

NOTE. Illustrative example of nominal (a and c) and disjunctive (b and d) coding of illustrative SNPs and diagnosis (Dx) plus *APOE*. For SNP 1, all genotypes have a sufficient frequency and are coded (à la genotypic model), but for SNP 2, the minor homozygote (CC) does not occur frequently enough and is thus combined with the heterozygote (à

la dominant model). In all tables, a 1 indicates the presence of a particular level of a categorical variable while a 0 indicates absence (e.g., Subj.1 is an Alzheimer's Disease patient, with 2 *APOE* E4 alleles, the AG genotype for SNP1 and either a CC or CA [presence of minor allele] for SNP2). Note that one subject has missing data (i.e., "*<NA>*"). This subject's data for SNP 1 is imputed to the mean of the sample for that SNP (SNP 1) where AA occurs in 70% of the sample, AG in 20%, and GG in 10%, therefore the missing data are imputed to those values.

MCA was used to analyze disjunctive tables (see Table 2). MCA is analogous to principal components analysis (PCA) tailored for categorical data. Like PCA, MCA uses the singular value decomposition and produces orthogonal components that are rank-ordered by their explained variance. Also like principal components analysis, MCA produces component (a.k.a. factor) scores for individuals (rows) and variables (columns). We used MCA instead of PCA to correct for stratification effects (in the preprocessing steps).

The next set of techniques we used were specific forms of PLS-CA. In general, PLS-CA is the generalization of partial least squares—a family of techniques that analyze the information common to two data tables [50]—that can be used with virtually any data type [49,60]. We used discriminant and seed PLS-CA in our studies here. Discriminant PLS-CA maximizes the separation between *a priori* groups of participants (see also [56]). Seed PLS-CA borrowed its name from the functional neuroimaging literature (e.g., [61–63]; see also [50]), where a seed (specific region or regions of interest) from the brain is selected as one data set and the remaining data are the second data set. In brain imaging, seed analyses look for brain regions correlated to the seed, here seed analysis look for genes with distributions of genotypes similar to the seed (i.e., high linkage disequilibrium). We used discriminant PLS-CA to maximally separate the diagnostic groups (i.e., CON, MCI, AD) and also find the genotypes most associated with each group. We used seed PLS-CA to identify genotypes with distributions similar to *APOE* (which is the strongest known genetic risk factor for non-familial, sporadic AD).

SNP Preprocessing

For all analyses we excluded any SNPs in the X and Y chromosomes, the pseudoautosomal region (XY), and mitochondrial region (i.e., we only analyzed SNPs in Chromosomes 1–22). All SNPs—which were genome-wide in the discovery study, and a much smaller subset in the validation study—were preprocessed with `PLINK` (v1.07; [64]) and in-house `R` code. SNP annotation was performed with the `NCBI2R` package [65], and the `biomaRt` package [66,67]. Regardless of study, SNPs were pre-processed with pipeline described below. Any specific information required (e.g., results from preprocessing parameters) is reported for each study in their respective sections.

We used the following quality control (QC) criteria: Participant and SNP call rates (i.e., completeness of data) ≥ 90%, minor allele frequency ≥ 5%, Hardy-Weinberg equilibrium $p \leq 10^{-6}$. SNPs were then recoded into a disjunctive format as seen in Table 2 (for more details on disjunctive format and PLS-CA see [49]). Some genotypes could occur below a reasonable threshold (e.g., 5%) and therefore we performed an additional preprocessing step: *genotypes* that fell below a 5% threshold were combined with another genotype. In our studies, only minor homozygotes were below the 5% threshold, and were thus combined with the heterozygotes which dichotomized the presence or absence of a minor allele (i.e., AA vs. {Aa or aa}; analogous to a dominant model). Missing genotypes were imputed to the mean of the sample in disjunctive format (i.e., assigned a probability of genotype occurrence; see Table 2 for an example). MCA was then applied to the participant × disjunctive genotypes matrix to detect and then remove stratification effects of sex, (2) ethnicity and/or race (à la PCA, multidimensional scaling, or other eigen stratification techniques).

Study design and overview

We conducted a two-part study: discovery and validation. In the discovery phase there were two analyses with ADNI-1 genome-wide SNPs. The results from the discovery analyses in ADNI-1 were used to create candidate SNP panels for validation in ADNI-GO/2. In the validation phase there was one analysis with a specific subset of ADNI-GO/2 SNPs.

Data from ADNI-1 and ADNI-GO/2 were *not* combined or preprocessed together at any stage in this study. We kept the two data sets completely separate 1) to ensure independence of the two data sets and also to ensure that no contamination occurred from or influence of one set on the other and 2) to guarantee the independence required for a proper discovery-validation pipeline (i.e., replication). Because ADNI-1 and ADNI-GO/2 have two different—and not entirely overlapping—chip sets we generated a candidate panel of SNPs for ADNI-GO/2 based on the SNPs and their associated genes identified in ADNI-1 (discovery). We describe all background required for preprocessing and analyses for discovery, candidate panel creation, and validation in the next sections as well as in their respective *Results* subsections.

Discovery study preprocessing and analyses

Discovery analyses were conducted on the ADNI-1 data set. Once data were preprocessed, we performed two types of PLS-CA: discriminant and seed PLS-CA. Discriminant PLS-CA was performed on diagnostic group (AD, MCI, CN) × genotypes. The goal of the Dx × genotype analyses was to detect genotypes most associated with each diagnostic category. The seed PLS-CA used *APOE* E4 alleles as the seed, and thus was performed on *APOE* (0, 1, or 2 E4 alleles) × genotypes. The goal of the *APOE* × genotype analyses was to identify new candidate genotypes with a distribution similar to *APOE* (i.e., high linkage disequilibrium). Both analyses identify candidate markers of AD: the discriminant (diagnosis-based) analysis identifies genotypes most associated with each group, whereas the seed analysis identifies genotypes

similar to *APOE* (i.e., high linkage disequilibrium), which we refer to as "Dx-GWAS" (i.e., diagnosis-based GWAS) and "*APOE*-GWAS" (i.e., *APOE*-based GWAS), respectively. All PLS-CA analyses used bootstrap resampling [68] to identify genotypes that are stable under resampling. The distributions around the genotype—via bootstrapping—can be tested with "bootstrap ratios" (BSR; [49–51]) which are analogous to *t*- or *Z*-tests and thus indicate significant and stable genotypes. Significant genotypes in our two GWAS (in ADNI-1) were then used to create a new candidate panel of SNPs for validation (in ADNI-GO/2).

Creation of SNP panels from Discovery for Validation

Because ADNI-1 and ADNI-GO/2 were used as independent (and thus isolated) data sets in our study, and because the data come from two different genome-wide chips, we used the significant genotypes from the discovery analyses (i.e., Dx-GWAS and the *APOE*-GWAS) to generate candidate SNPs for validation. We did so with the following procedure. For all significant genotypes in the discovery analyses we used their respective SNPs to: (1) compile a list of all SNPs within a 50kbase (25+/-) window of those SNPs, and (2) retrieve all stable ensembl gene (ENSG) IDs associated with those SNPs from the discovery analyses and in turn retrieved *all possible SNPs* associated those ENSG IDs. Both steps were performed with `biomaRt` and `NCBI2R`. All SNPs from the steps (1) and (2) were then aggregated. We then extracted all SNPs from the ADNI-GO/2 data that were from the discovery-derived aggregate set. These extracted SNPs were then used in the validation phase.

Validation study preprocessing and analyses

Validation analyses were conducted on the ADNI-GO/2 data set. Once the data were preprocessed, we performed one final PLS-CA: a discriminant PLS-CA on the validation SNPs

based on diagnosis. As in the discovery phase, the goal of the validation Dx × genotype analysis was to detect genotypes most associated with each diagnostic category. However, in the validation set we have a smaller and restricted set of SNPs in the candidate panel (derived from the discovery analyses). As in the discovery phase, we used bootstrap resampling and the BSR test to identify genotypes stable under resampling.

Results

The goal of discovery phase was to identify potential genotypic candidates of AD and CON through two analyses conducted on the ADNI-1 data set: (1) a GWAS to identify genotypes most associated with Dx and (2) a GWAS to identify genotypes in high linkage disequilibrium with *APOE*. Results from discovery analyses were used to generate a candidate set of SNPs for further analyses in a validation phase in ADNI-GO/2. For all PLS-CA component maps we use what is called the "asymmetric plot" of component scores.

Discovery (ADNI-1)

ADNI-1 genome-wide data contains 620,901 SNPs and 757 participants. After QC and preprocessing 756 participants (AD = 344, MCI = 204, CON = 208) and 517,441 SNPs (in chromosomes 1–22) remained. The data matrix we analyzed was 756 participants × 1,332,455 genotypes. MCA was applied to the 756 × 1,332,455 matrix to identify confounding effects. The first two components displayed a strong effect of race, a minor effect of ethnicity, and no effect of sex and were thus removed from (i.e., regressed out of) the data. Subsequent components showed no apparent effects of stratification.

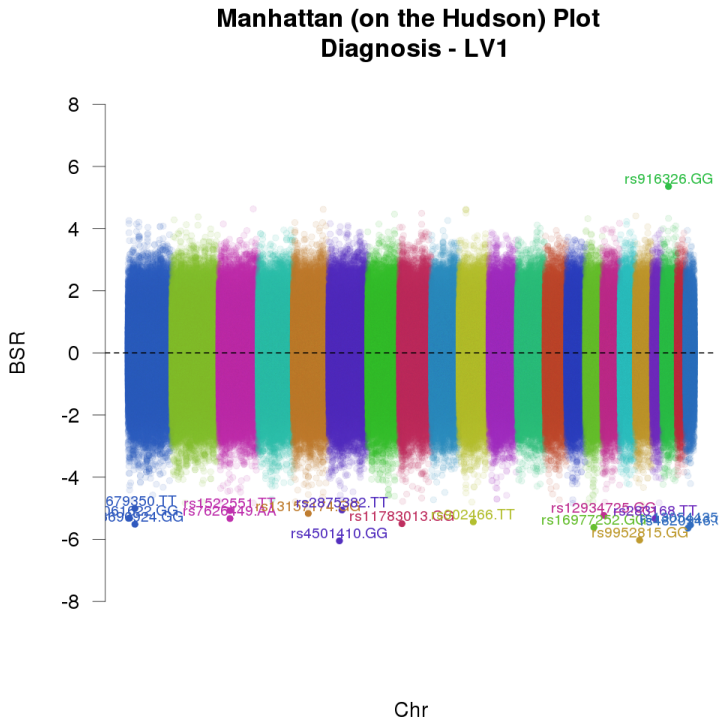
Two discriminant PLS-CAs were performed: one on Dx × genotypes and another on *APOE* × genotypes. We describe each analysis in turn. For the discovery GWAS, we used a

cutoff of ±5 for the BSR tests which is slightly below the traditional GWAS significance threshold ($p < .05 \times 10^{-8}$ would correspond to a BSR ≈ 5.33).

Dx-GWAS. Because there were only three groups (AD, MCI, and CON), the discriminant PLS-CA produced only two components. Component 1 explained 50.25% and was driven by the separation of the AD group from the MCI group (see Supplemental Figure 1). Component 2 explained 49.76% of the variance and separates the CON group from the other two groups (see Supplemental Figure 1). Supplemental Figure 1 shows the distribution of all significant genotypes (based on BSRs) on the component map, with respect to the groups. Figure 1 shows all genotypes plotted with their BSR values in a Manhattan-like plot. BSR values can be positive or negative—because the sign matches their component score—we call this plot "Manhattan on the Hudson" (MotH; like a city skyline on a river). Figure 1a shows the BSRs for all genotypes for Dx-GWAS Component 1 and Figure 1b shows the BSRs for all genotypes for Dx-GWAS Component 2. In the component map, most significant genotypes appear in the upper left quadrant (Supplemental Figure 1), and are thus more related to the AD group than the other groups. Significant genotypes, and additional information about their SNPs, are listed in Table S1a for Component 1 and Table S1b for Component 2.

APOE-GWAS. Because there were only 3 groups (respectively, "0 E4," "1 E4," and "2 E4" alleles), seed PLS-CA produced only two components. Component 1 explained 51.24% of the total variance and is driven by the presence (left) vs. absence (right) of E4 alleles (see Supplemental Figure 2). Component 2 explained 48.76% of the variance and separates the two E4 alleles group from the other two groups (see Supplemental Figure 2). Supplemental Figure 2 shows the distribution of all significant genotypes on the component map, with respect to the groups. Note that rs2075650 genotypes map almost exactly onto the *APOE* categories; alone

rs2075650 is an almost perfect proxy of *APOE* genotype. Figure 2 shows the BSRs for all genotypes from this analysis in a "Manhattan on the Hudson" plot. Figure 2a shows the BSRs for all genotypes for *APOE*-GWAS Component 1 and Figure 2b shows the BSRs for all genotypes for *APOE*-GWAS Component 2. In the component map, the most significant genotypes appear in the upper part of Component 2 (Supplemental Figure 2). Significant genotypes, and additional information about their SNPs, are listed in Table S2a for Component 1 and Table S2b for Component 2 (and genotypes significant on both components are listed in Table S2c).



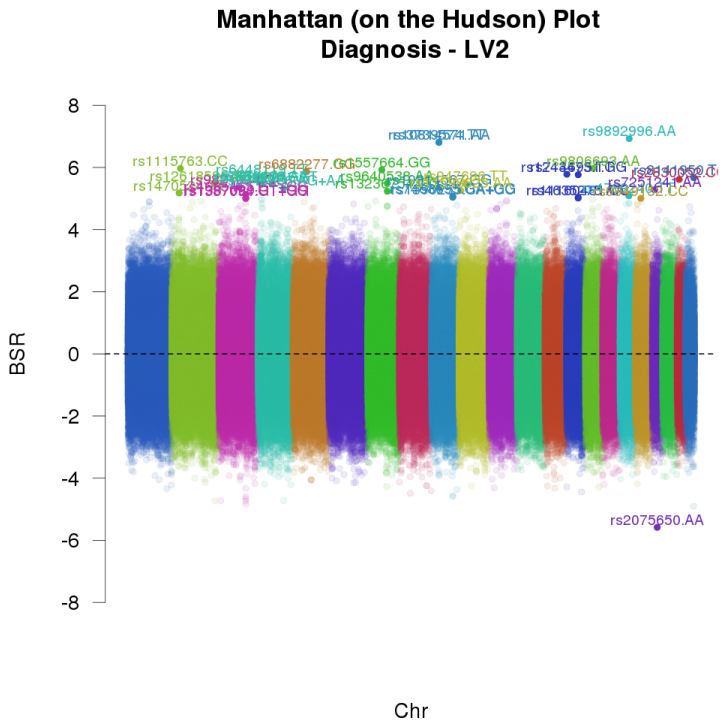


Figure 1. Manhattan (on the Hudson) plots for the multivariate Diagnosis (Dx)-GWAS (via DiCA). Horizontal axes are each genotype ordered by Chromosome (Chr), where each Chr is color-coded (1–22). The vertical axes are bootstrap ratio values (BSRs). Both panels show BSRs (analogous to *t*- or *Z*-scores; which can be positive or negative) for each genotype along Component 1 (a.k.a. Latent Variable (LV) 1; panel a, top) and Component 2 (a.k.a. Latent Variable (LV) 2; panel b, bottom) – the same components as in Figure 2. With respect to the multivariate Dx-GWAS a wide variety of genotypes show significant, and similar, effects and are not concentrated in any particular region (see also Table S1 and Figure S1).

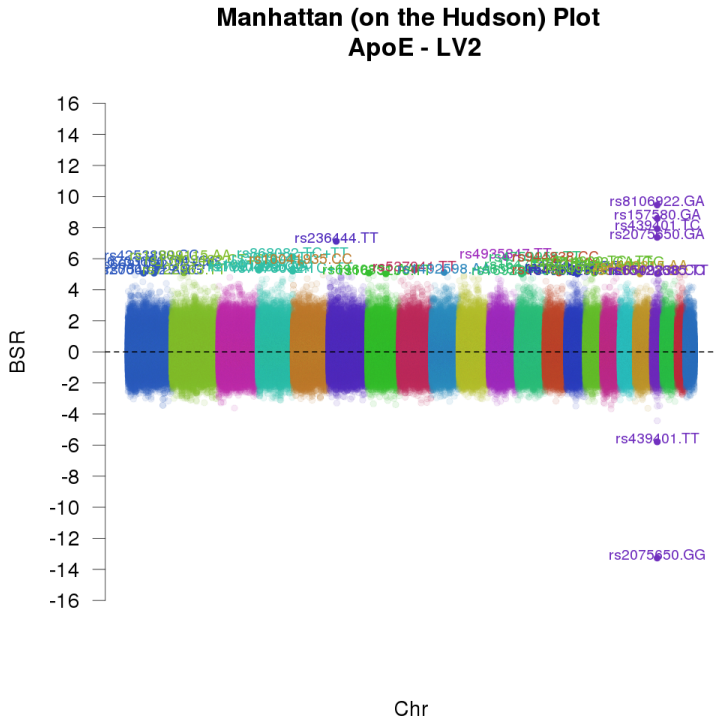


Figure 2. Manhattan (on the Hudson) plots for the multivariate *APOE*-GWAS (via DiCA). Horizontal axes are each genotype ordered by Chromosome (Chr), where each Chr is color-coded (1–22). The vertical axis is bootstrap ratio values (BSRs). Both panels show BSRs (analogous to *t*- or *Z*-scores; which can be positive or negative) for each genotype along Component 1 (a.k.a. Latent Variable (LV) 1; panel a, left) and Component 2 (a.k.a. Latent Variable (LV) 2; panel b, right) – the same components as in Figure 4. With respect to the multivariate *APOE*-GWAS, many

of the effects are concentrated, generally, in Chr19 (see also Table S2) across both components, but much more so for Component 1 (a; left). While Component 2 (b; right) shows a variety of effects, some of the strongest are still in Chr19.

Candidate panel creation

All SNPs associated with significant genotypes were used for creation of the candidate panel, regardless of the effects with which there were most associated (e.g., CON, or 2 E4 *APOE* alleles). A total of 105 genotypes from 96 SNPs exceeded the ±5 BSR threshold (see Supplemental Tables 1 and 2). From these 96 SNPs, we created a candidate panel of SNPs in two steps: (1) we identified all SNPs within a 50kbase (25+/–) window of these 96 SNPs, and (2) we retrieved all stable ENSG IDs associated with these 96 SNPs and in turn retrieved all possible SNPs associated with stable ENSG IDs. From step (1) there were 232,120 possible SNPs, and from step (2) there were 956,013 SNPs from 72 ENSG IDs. Steps (1) and (2) identified 1,045,360 possible SNPs. From this large set we extracted all available SNPs in the ADNI-GO/2 chip (see next section).

Validation (ADNI-GO/2)

ADNI-GO/2 genome-wide data contains 730,525 SNPs and 791 participants. From the candidate panel, we extracted 5,508 SNPs from the 730,525 available SNPs and performed the QC and preprocessing steps. After QC and preprocessing 791 participants (AD = 203, MCI = 319, CON = 269) and 5,508 SNPs remained. The data matrix we analyzed was 791 participants × 14,200 genotypes. MCA was applied to the 791 × 14,200 matrix to identify confounding effects. The first two components had a strong effect of race, a minor effect of ethnicity, and no effect of sex and were thus removed from (i.e., regressed out of) the data. Subsequent components showed no apparent effects of stratification.

A discriminant PLS-CA (Dx × genotypes) was performed on the 791 × 14,200 matrix. For the validation analysis, we used a cutoff of ±3.25 for the BSR tests. This value corresponds to a Bonferroni cutoff for the number of unique ENSG IDs tested. Because there were only 3 groups (AD, MCI, and CON), the discriminant PLS-CA produced only two components. Component 1 explained 50.41% and was driven by the separation of the MCI group from the AD and CON groups (see Figure 3). Component 2 explained 49.59% of the variance and separated the CON group from the AD group (see Figure 3).

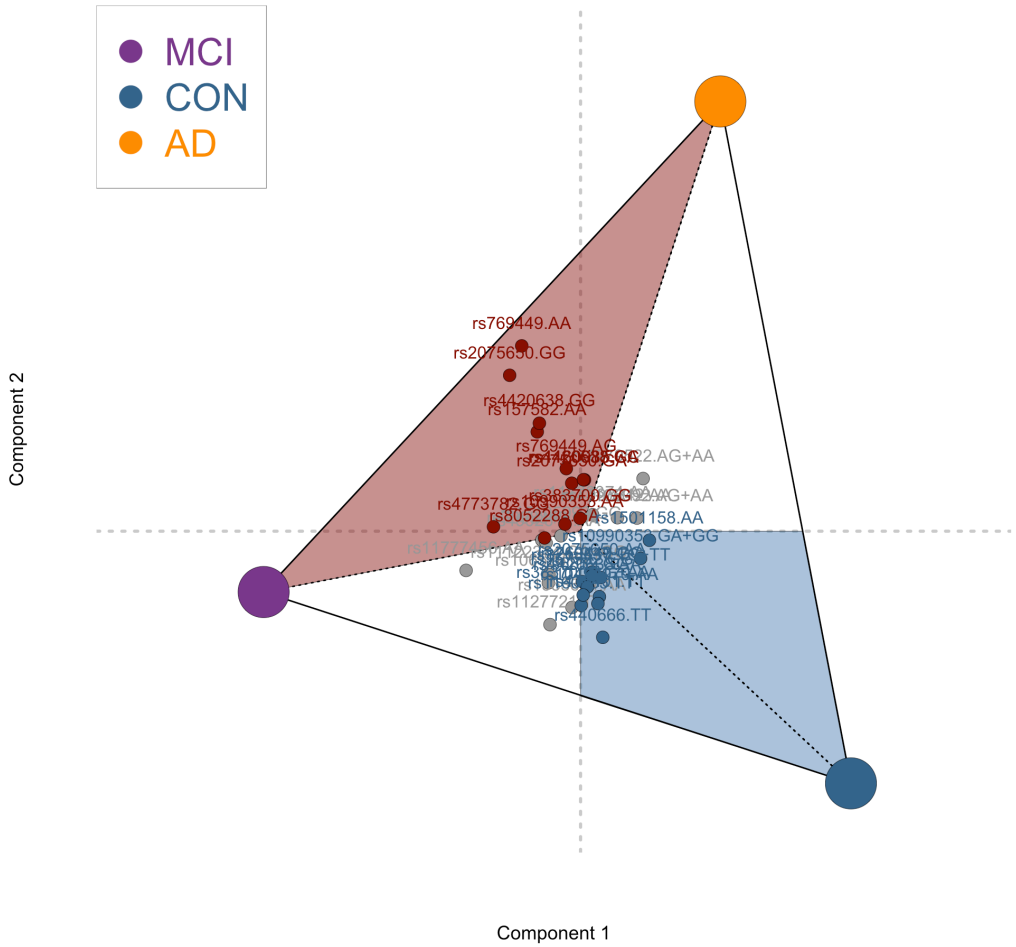


Figure 3. Diagnosis-based DiCA in the Validation phase. The groups form a boundary region on the components. We denoted portions of this subspace as “control” associated (blue) or “risk” associated (red); anything outside the

control or risk regions is ambiguous. Genotypes that fall in the “control” region are more associated with the control group than the other groups and genotypes that fall in the “risk” region are more associated with {MCI or AD} than the CON group.

Table 3

Validation set: significant genotypes on either component

rsid	genotype	BSR - 1	BSR - 2	chr	Gene symbol	Risk / Control Region
rs10494979	AA	0.95	*-3.422*	1	*PTPN14*	CON
rs11122374	AA	1.985	*3.255*	1	*TSNAX-DISC1,DISC1*	
rs11122374	GA+GG	-1.995	*-3.346*	1	*TSNAX-DISC1,DISC1*	
rs1501158	AA	*3.319*	-0.409	4	*LINC00504*	CON
rs1065261	TT	-1.971	*-3.451*	6	*CPNE5*	
rs3213537	TC+TT	1.345	*-3.31*	6	*CPNE5*	CON
rs11777456	AA	*-3.366*	-1.615	8	*NCALD*	
rs10990353	AA	*-3.374*	1.483	9	*ZYG11AP1 / LOC100421294*	RISK
rs10990353	GA+GG	*3.496*	-1.482	9	*ZYG11AP1 / LOC100421294*	CON
rs7093342	GA	*-3.395*	-0.848	10	*ITIH5*	
rs7093342	AA	*3.904*	1.5	10	*ITIH5*	
rs1979522	AG+AA	*3.275*	2.298	12	*LRMP*	
rs4773782	GG	*-3.608*	0.185	13	*GPC6*	RISK
rs17105992	GG	*-3.513*	-0.847	14	*LOC107984016,RAD51B*	
rs17105992	AG+AA	*3.55*	0.844	14	*LOC107984016,RAD51B*	
rs4902611	AA	*-3.93*	-1.293	14	*LOC107984016,RAD51B*	
rs8052288	GA	*-3.476*	-0.692	16	*RBFOX1*	RISK
rs1553614	AA	-0.278	*-3.272*	16	*RBFOX1*	
rs6859	GG	0.415	*-3.895*	19	*NECTIN2**	CON
rs2075650	GA	-0.656	*3.5*	19	*TOMM40*	RISK
rs2075650	AA	1.555	*-5.558*	19	*TOMM40*	CON
rs2075650	GG	-2.12	*4.338*	19	*TOMM40*	RISK
rs157582	GG	0.162	*-5.297*	19	*TOMM40*	CON
rs157582	AA	-1.773	*3.72*	19	*TOMM40*	RISK
rs1160985	CC	0.344	*4.349*	19	*TOMM40*	RISK
rs1160985	TT	0.048	*-4.101*	19	*TOMM40*	CON
rs769449	GG	1.749	*-6.527*	19	*APOE*	CON
rs769449	AA	-1.792	*5.195*	19	*APOE*	RISK
rs769449	AG	-0.995	*4.194*	19	*APOE*	RISK
rs439401	TT	0.678	*-3.277*	19		CON
rs4420638	GA	0.207	*3.95*	19	*APOC1*	RISK
rs4420638	AA	0.803	*-6.458*	19	*APOC1*	CON
rs4420638	GG	-1.637	*3.913*	19	*APOC1*	RISK

rs383700	GG	-0.179	*3.314*	21	*APP*	RISK
rs383700	AG+AA	0.179	*-3.436*	21	*APP*	CON
rs440666	TT	0.631	*-3.88*	21	*APP*	CON
rs1127721	TT	-0.789	*-3.543*	22	*PARVG*	

NOTE. * rs6859 previously associated with *PVRL2*. Significant genotypes for both components from the validation set. Component 1 primarily separated CON from AD, whereas Component 2 primarily separated AD from the other two groups. In the validation analysis, we used the asymmetric version of correspondence analysis (CA) to help visualize which genotypes are most associated with each group, and to create identify "control" (CON) and "risk" (RISK) regions within the CA components (see Figure 3). Genotypes labeled as CON or RISK indicate which region they are in, otherwise their association is more ambiguous in terms of association with risk or control ("protective") factors.

All significant genotypes from the validation analysis are found in Table 3. In figure 3, we used the triangle (i.e., the "simplex" from CA) whose vertices were generated by the groups and where the component axes (horizontal and vertical lines) are plotted to create boundaries in order to interpret genotypic effects. We focused specifically on two boundaries: (1) the boundary associated mostly with the CON group (lower right, blue) and (2) the boundary associated with "risk status" (i.e., AD or MCI; upper middle to middle left, red). The relevant SNPs are highlighted in their boundary colors (blue for CON; red for "risk status") in Figure 3, and identified as part of the "RISK" or "CON" region in Table 3.

Discussion

Recently there has been an increased interest in multivariate approaches for the genetics of AD [69]. However many of these studies in AD still only use the additive model contrary to expected non-additive effects and work that illustrated that the genotypic [48] or co-dominant [47] models are better suited for complex effects. Furthermore, multivariate approaches—also based on correspondence analysis—that emphasize the genotypic model have been used extensively to understand populations in animals (see, e.g., [70]) and—albeit rarely—some

studies in humans [71–75]. Note that throughout the discussion we denote gene names in fully capital and italicized font.

Our newly developed technique (PLS-CA) gave us a much more specific view of the genetics of AD than other studies: with emphasis on the genotypes instead of to the SNPs. We found a wide array of genotypes from SNPs (and their respective genes) across both *Discovery* and *Validation*. We provide only a brief discussion of the *Discovery* phase itself, with emphasis on particularly interesting effects and how the *Discovery* phase leads to the *Validation* phase. Our discussion emphasizes the results from the *Validation* phase and in particular the effects we have identified "risk" or "control" associated effects (see Fig. 3 and Table 3).

Discovery

The aim of *Discovery* was to identify the genotypes that express patterns that match, as closely as possible, the (1) diagnostic groups (i.e., "Dx-GWAS") or (2) *APOE* E4 distribution (i.e., "*APOE*-GWAS"). Both analyses aimed to reveal new candidate genotypes from two different perspectives: top down (Dx) and bottom up (*APOE*).

The *APOE*-GWAS revealed fewer effects than the Dx-GWAS but these effects were almost uniformly stronger than the Dx-GWAS and were generally in high disequilibrium with the presences vs. absence of *APOE* E4 (see Fig. 2, Supplemental Fig. 2, and Supplemental Table 2a and 2c). There were two striking effects in the *APOE*-GWAS: First, the *TOMM40* genotypes for rs2075650 express an pattern almost identical to *APOE.* This suggests that rs2075650 has almost the same predictive power as, and is in almost perfect linkage disequilibrium with, the two biallelic SNPs used to genotype for *APOE*. While the relationship between *APOE* and *TOMM40* is not surprising [23,76] the effect we observed (i.e., near perfect disequilibrium with

APOE) was surprising. The second striking effect was that the GG genotype of rs10830213 in the *RAB38* gene in Chromosome 11 was one of the strongest non-Chromosome 19 effects and GG genotype of rs10830213 was strongly associated with the absence of *APOE* E4 and thus suggested a protective effect (in ADNI-1). The *RAB38* gene has recently been associated with frontotemporal dementia [77], whereas other *RAB*-family genes have been associated with multiple neurodegenerative disorders [78]. *RAB38* is of particular interest because it has also been associated with the production of Aβ [79].

The Dx-GWAS revealed a very wide array of effects across the genome (see Fig. 1, Supplemental Fig. 1 and Supplemental Table 1). The wide array of genotypic effects suggests that very many genes are specifically associated with AD, MCI, or CON groups within ADNI-1. Some example effects include contributions of: *EFNA5* which is broadly associated with AD [80] and hippocampal atrophy in AD [40]. Interestingly and somewhat contrary to our expectations the "*APOE*-GWAS" did not detect any genotypes specifically associated with *APP*—a well-known causal factor to some forms of AD [81,82]—rather, effects of *APP* were observed in the "Dx-GWAS".

Validation

While there were a number of very interesting effects in *Discovery*, the *Validation* phase was critical to identify the most likely reproducible effects. Our *Validation* phase (ADNI-GO/2) was similar to the Dx-GWAS in *Discovery*, in that we aimed to identify genotypes most associated with diagnosis. However, the set of genotypes for *Validation* were derived from a candidate panel creation process (see *Methods*) based on the significant markers from the Dx-GWAS and the *APOE*-GWAS. Our discussion of the *Validation* phase is focused on the

genotypes we have identified as “risk” or “control” associated effects, though we do provide some discussion of significant genotypes that are, from our analyses, ambiguous or undetermined in how they contribute to the overall effects.

Chromosome 19. Some of the strongest effects in our validation analyses were associated with genotypes in Chromosome 19 (Chr19). Specifically, there were strong effects in *APOE* or strong effects in disequilibrium with *APOE*, such as *APOC*, and *NECTIN2* (by way of a SNP previously associated with *PVRL2*; see Table 3). For AD, the effects of *APOE* and Chr19 in general are fairly well known. However, our results for these genotypes are important for two reasons: (1) these findings support our development and use of PLS-CA, as they found some of the most well-known effects and identified them as the strongest sources, and (2) our approach provides a specificity of the effects beyond the usual SNP-based analyses; specifically, we have identified exactly which genotypes contribute to these effects and we can characterize the genotypes as “risk” or “control” associated (see Fig. 3 and Table 3). For examples, the ‘AA’ genotype of rs769449 (*APOE*) and the ‘GG’ genotype of rs2075650 (*TOMM40*) are at the extreme of our asymmetric plot almost exactly half way between AD and MCI and directly opposite of the CON group (Fig. 3). This means that AA in rs769449 and GG in rs2075650 rarely occur in CON, and tend to occur roughly equally in both AD and MCI. Furthermore, and most importantly for Chr19 effects: our *Discovery* phase identified ‘AA’ of rs6859 associated with the *presence* of *APOE* E4, whereas our *Validation* phase identified ‘GG’ of rs6859 associated with the CON group. While the “direction” of the effect is the same across both studies, the source of the effect (i.e., minor allele vs. major allele) differs across both studies and may reflect, to some degree, the make-up of each study (i.e., because ADNI-1 is more risk prone, where as ADNI-GO/2 is not).

APP. As previously noted, *APP* was identified in the Dx-GWAS—not the *APOE*-GWAS—in *Discovery*. *APP* markers continued to show an effect in the validation analyses (see Table 3). However, like with some of the markers in Chromosome 19, we find a nuanced story about the genotypes associated with *APP*. In *Validation*, we see *APP* markers associated both with "risk" and "control" (Fig. 3 and Table 3). The same kind of effect was observed in rs440666: 'TT' was associated with "control" in our validation study (Table 3 and Fig. 3). Given the findings across both our Discovery and Validation phases, our findings here suggest that *APP* is more related to diagnostic criteria than to *APOE* distribution, and that specific *APP* genotypes are likely *protective*.

RBFOX1 and GPC6. In our analyses, we found risk specific genotypes in the *GPC6* and *RBFOX1* genes (Table 3; Fig. 3). These two genes are not commonly identified in Alzheimer's or in other dementias. However, contributions of both of these genes have been reported very recently. *RBFOX1* has been associated with neurofibrillary tangles in AD [83]. *RBFOX1* has also been associated (structural) imaging phenotypes [84] and hippocampal volume [85]. *GPC6* has been associated with AD at the whole genome scale [86] as well as the decline in ADAS-cog measure [87].

Novel genetic risk and "protective" factors. Our analyses revealed several novel genetic contributions to either risk or control (possibly "protective") associated effects. Those effects are: *PTPN14* ("control"), *CPNE5* ("control"), *LINC00504* ("control"), and intergenic between *ZYG11AP1* and *LOC100421294* (both "control" and "risk"). As of now, there is no reported effect of rs1501158 (*LINC00504*) and rs10990353 (intergenic, *ZYG11AP1* / *LOC100421294*) in AD, other dementias, or in general the cognitive aging literature. Additionally, neither *CPNE5* nor *PTPN14* have been reported in AD but have been reported in other domains. However,

CPNE5 has been associated with a variety of other neurodegenerative diseases [88] and with *CPNE5* expression levels in the frontal cortex with respect to aging [89]. Furthermore, brain-specific variants of *PTPN14* have been reported in AD (postmortem; [90]). Finally, *PTPN14* is known to interact with *KIBRA* [91], which is a gene often implicated in memory [92].

Stable but ambiguous effects. Finally, there were a number of significant and stable effects through both the discovery and validation analyses. In the validation analysis, these effects were "ambiguous" in that we have not classified them as "risk" or "control" factors because they fall outside of our designated regions (i.e., ambiguous because effects were either in a region between CON and AD or between CON and MCI). *LRMP* has been reported both in Huntington's Disease [93] and with substantial down-regulation in HIV neuroblastoma cell lines [94]. *PARVG* has been associated with Parkinson's Disease [95] and neurodegeneration [96]. Though *ITIH5* has no reported effects in the AD literature, it is amongst a small number of candidates for CSF in AD [97]. While the remaining genes—*RAD51B*, *NCALD*, and *DISC1*—are rarely reported in AD, they share a common trait: they have been associated with the amyloid precursor gene *APP* [98,99]. Furthermore, it has been suggested that *DISC1* plays a much broader role than "disrupted in schizophrenia" for *DISC1* [100]. *DISC1* has been reported in late-onset AD [101] and has also been associated with both *APP* and Aβ production [102]. Furthermore, *DISC1* plays a broad role in a common phenotype across many neurological and psychiatry disorders: cortical thickness in many brain regions [103]. Finally, *RAD51B* has been associated with macular degeneration [104,105].

Limitations

While our study presents a first of its kind analysis that reveals both novel genetic contributions and more specificity of known genetic effects in AD there are some limitations. First, the ADNI data sets are, by today's standards, relatively small samples for such a study. However, our *Discovery* and *Validation* sets were kept completely separate from one another (i.e., never combined and no shared preprocessing). That being said an optimal study would include a third completely independent data set to explicitly test the sensitivity and specificity of diagnosis based on the genotypes we discovered; this could be done either with a simple coding of risk vs. non-risk genotypes or by using weighted genotypes (via our bootstrap ratios) which would be akin to a polygenic risk score. Additionally, in our *Validation* phase we could have weighted the candidates based on aggregate bootstrap ratios from the *Discovery* phase. However, we specifically chose not to and opted for a more data-driven strategy. Ultimately this was beneficial: for example, if we had used weighted values in the *Validation* phase we may have missed the effects of rs6859. Recall that different genotypes of rs6859 expressed different effects in *Discovery* compared to *Validation*.

Conclusions

Like in our work here [84–87] use the ADNI sample. However, all of these papers, as well as [83] all use more complex study designs or more complex methods, which in some cases [84,85] demand vast computational resources. Other work, such as [87], also have more data and larger samples. Yet our much simpler and computationally less expensive strategies with PLS-CA found many of the same effects reported in all of the aforementioned papers, such as the contributions of *GPC6* and *RBFOX1*.

Furthermore, PLS-CA found a number of long-studied and already well-known genetic

effects in AD such as the contributions across Chr. 19 and the strong disequilibrium of *APOE-TOMM40*. While these findings are not surprising, it is important that new methods—such as ours—identify well-known and robust effects in order to illustrate validity and appropriateness for the problems at hand.

We had initially developed PLS-CA specifically to handle the complexities of genetics in AD (see [49]). With our application of PLS-CA here it is clear that PLS-CA has many clear advantages over traditional and more recent approaches to GWAS. First, PLS-CA does not make assumptions about genotypic effects; rather, PLS-CA will reveal what type of genotypic effects exists (e.g., dominant vs. additive) and the directions of these effects. More importantly, another major advantage PLS-CA has over traditional approaches is that PLS-CA is a multivariate technique and thus actually suited for discovery of polygenic effects. Together these features will help to more reliably identify complex effects and thus hopefully reduce false positives and non-replications in the literature.

References

[1] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–3. doi:10.1126/science.8346443.

[2] Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 1994;7:180–4. doi:10.1038/ng0694-180.

[3] Mahley RW. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. Science 1988;240:622–30. doi:10.1126/science.3283935.

[4] Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet 2007;39:168–77. doi:10.1038/ng1943.

[5] Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 2011;43:436–41. doi:10.1038/ng.801.

[6] Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009;41:1088–93. doi:10.1038/ng.440.

[7] Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 2009;41:1094–9. doi:10.1038/ng.439.

[8] Mez J, Chung J, Jun G, Kriegel J, Bourlas AP, Sherva R, et al. Two novel loci, COBL and SLC10A2, for Alzheimer's disease in African Americans. Alzheimers Dement 2017;13:119–29. doi:10.1016/j.jalz.2016.09.002.

[9] Jun GR, Chung J, Mez J, Barber R, Beecham GW, Bennett DA, et al. Transethnic genome-wide scan identifies novel Alzheimer's disease loci. Alzheimers Dement 2017;13:727–38. doi:10.1016/j.jalz.2016.12.012.

[10] Bertram L, Tanzi RE. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. Nat Rev Neurosci 2008;9:768–78. doi:10.1038/nrn2494.

[11] Karch CM, Cruchaga C, Goate AM. Alzheimer's Disease Genetics: From the Bench to the Clinic. Neuron 2014;83:11–26. doi:10.1016/j.neuron.2014.05.041.

[12] Schellenberg GD, Montine TJ. The genetics and neuropathology of Alzheimer's disease. Acta Neuropathol (Berl) 2012;124:305–23. doi:10.1007/s00401-012-0996-2.

[13] Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 2014;505:550–4. doi:10.1038/nature12825.

[14] van der Lee SJ, Holstege H, Wong TH, Jakobsdottir J, Bis JC, Chouraki V, et al. PLD3 variants in population studies. Nature 2015;520:E2–3. doi:10.1038/nature14038.

[15] Lambert J-C, Grenier-Boley B, Bellenguez C, Pasquier F, Campion D, Dartigues J-F, et al. PLD3 and sporadic Alzheimer's disease risk. Nature 2015;520:E1–E1. doi:10.1038/nature14036.

[16] Okuizumi K, Tsuji S. Alzheimer's disease as a polygenic disease. Neuropathology 1998;18:111–5. doi:10.1111/j.1440-1789.1998.tb00087.x.

[17] Desikan RS, Fan CC, Wang Y, Schork AJ, Cabral HJ, Cupples LA, et al. Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. PLOS Med 2017;14:e1002258. doi:10.1371/journal.pmed.1002258.

[18] Escott-Price V, Myers AJ, Huentelman M, Hardy J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. Ann Neurol 2017;82:311–4. doi:10.1002/ana.24999.

[19] Hohman TJ, Koran ME, Thornton-Wells T, for the Alzheimer's Neuroimaging Initiative. Epistatic Genetic Effects among Alzheimer's Candidate Genes. PLoS ONE 2013;8:e80839. doi:10.1371/journal.pone.0080839.

[20] Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivières S, Jahanshad N, et al. Common genetic variants influence human subcortical brain structures. Nature 2015;520:224–9. doi:10.1038/nature14101.

[21] Lehner B. Molecular mechanisms of epistasis within and between genes. Trends Genet 2011;27:323–31. doi:10.1016/j.tig.2011.05.007.

[22] Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O, et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. Mol Psychiatry 2011;16:903–7. doi:10.1038/mp.2011.52.

[23] Roses AD, Lutz MW, Amrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS, et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. Pharmacogenomics J 2010;10:375–84. doi:10.1038/tpj.2009.69.

[24] Lautner R, Palmqvist S, Mattsson N, Andreasson U, Wallin A, P\a alsson E, et al. Apolipoprotein E Genotype and the Diagnostic Accuracy of Cerebrospinal Fluid Biomarkers for Alzheimer Disease. JAMA Psychiatry 2014;71:1183–1191.

[25] Trabzuni D, Wray S, Vandrovcova J, Ramasamy A, Walker R, Smith C, et al. MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. Hum Mol Genet 2012;21:4094–103. doi:10.1093/hmg/dds238.

[26] Lancaster TM, Brindley LM, Tansey KE, Sims RC, Mantripragada K, Owen MJ, et al. Alzheimer's disease risk variant in CLU is associated with neural inefficiency in healthy individuals. Alzheimers Dement 2015;11:1144–52. doi:10.1016/j.jalz.2014.10.012.

[27] Ebbert MTW, Boehme KL, Wadsworth ME, Staley LA, Mukherjee S, Crane PK, et al. Interaction between variants in CLU and MS4A4E modulates Alzheimer's disease risk. Alzheimers Dement 2016;12:121–9. doi:10.1016/j.jalz.2015.08.163.

[28] Peterson D, Munger C, Crowley J, Corcoran C, Cruchaga C, Goate AM, et al. Variants in PPP3R1 and MAPT are associated with more rapid functional decline in Alzheimer's disease: The Cache County Dementia Progression Study. Alzheimers Dement 2014;10:366–71. doi:10.1016/j.jalz.2013.02.010.

[29] Lebedeva E, Stingl JC, Thal DR, Ghebremedhin E, Strauss J, Özer E, et al. Genetic variants in PSEN2 and correlation to CSF β-amyloid42 levels in AD. Neurobiol Aging 2012;33:201.e9-201.e18. doi:10.1016/j.neurobiolaging.2010.07.017.

[30] Naj AC, Jun G, Reitz C, Kunkle BW, Perry W, Park Y, et al. Age-at-Onset in Late Onset Alzheimer Disease is Modified by Multiple Genetic Loci. JAMA Neurol 2014;71:1394–404. doi:10.1001/jamaneurol.2014.1491.

[31] Cruchaga C, Kauwe JSK, Harari O, Jin SC, Cai Y, Karch CM, et al. GWAS of Cerebrospinal Fluid Tau Levels Identifies Risk Variants for Alzheimer's Disease. Neuron 2013;78:256–68. doi:10.1016/j.neuron.2013.02.026.

[32] Charney E. Genes, behavior, and behavior genetics. Wiley Interdiscip Rev Cogn Sci 2016:n/a-n/a. doi:10.1002/wcs.1405.

[33] Kauwe JSK, Cruchaga C, Karch CM, Sadler B, Lee M, Mayo K, et al. Fine Mapping of Genetic Variants in BIN1, CLU, CR1 and PICALM for Association with Cerebrospinal Fluid Biomarkers for Alzheimer's Disease. PLOS ONE 2011;6:e15918. doi:10.1371/journal.pone.0015918.

[34] Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. NeuroImage 2010;53:1051–63. doi:10.1016/j.neuroimage.2010.01.042.

[35] Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, et al. Voxelwise genome-wide association study (vGWAS). NeuroImage 2010;53:1160–74. doi:10.1016/j.neuroimage.2010.02.032.

[36] Hibar DP, Stein JL, Kohannim O, Jahanshad N, Saykin AJ, Shen L, et al. Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. Neuroimage 2011;56:1875–91. doi:10.1016/j.neuroimage.2011.03.077.

[37] Shulman JM, Chen K, Keenan BT, Chibnik LB, Fleisher A, Thiyyagura P, et al. Genetic susceptibility for Alzheimer disease neuritic plaque pathology. JAMA Neurol 2013;70:1150–1157.

[38] Swaminathan S, Shen L, Risacher SL, Yoder KK, West JD, Kim S, et al. Amyloid pathway-based candidate gene analysis of [11C]PiB-PET in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort. Brain Imaging Behav 2011;6:1–15. doi:10.1007/s11682-011-9136-1.

[39] Meda SA, Narayanan B, Liu J, Perrone-Bizzozero NI, Stevens MC, Calhoun VD, et al. A large scale multivariate parallel ICA method reveals novel imaging–genetic relationships for Alzheimer's disease in the ADNI cohort. NeuroImage 2012;60:1608–21. doi:10.1016/j.neuroimage.2011.12.076.

[40] Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, et al. Hippocampal Atrophy as a Quantitative Trait in a Genome-Wide Association Study Identifying Novel Susceptibility Genes for Alzheimer's Disease. PLoS ONE 2009;4:e6501. doi:10.1371/journal.pone.0006501.

[41] Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet 2006;7:781–91. doi:10.1038/nrg1916.

[42] Vormfelde SV, Brockmöller J. On the value of haplotype-based genotype–phenotype analysis and on data transformation in pharmacogenetics and -genomics. Nat Rev Genet 2007;8. doi:10.1038/nrg1916-c1.

[43] Balding DJ. Reply: On the value of haplotype-based genotype–phenotype analysis and on data transformation in pharmacogenetics and -genomics. Nat Rev Genet 2007;8. doi:10.1038/nrg1916-c2.

[44] Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al. Common variants in ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 2011;43:429–35. doi:10.1038/ng.803.

[45] Hooli BV, Lill CM, Mullin K, Qiao D, Lange C, Bertram L, et al. PLD3 gene variants and Alzheimer's disease. Nature 2015;520:E7–8. doi:10.1038/nature14040.

[46] Yan J, Aliev F, Webb BT, Kendler KS, Williamson VS, Edenberg HJ, et al. Using genetic information from candidate gene and genome-wide association studies in risk prediction for alcohol dependence. Addict Biol 2014;19:708–21. doi:10.1111/adb.12035.

[47] Lettre G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet Epidemiol 2007;31:358–62. doi:10.1002/gepi.20217.

[48] Monir MM, Zhu J. Comparing GWAS Results of Complex Traits Using Full Genetic Model and Additive Models for Revealing Genetic Architecture. Sci Rep 2017;7:38600. doi:10.1038/srep38600.

[49] Beaton D, Dunlop J, Abdi H, Alzheimer's Disease Neuroimaging Initiative. Partial Least Squares Correspondence Analysis: A Framework to Simultaneously Analyze Behavioral and Genetic Data. Psychol Methods 2016;21:621–51. doi:10.1037/met0000053.

[50] Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. NeuroImage 2011;56:455–75. doi:DOI: 10.1016/j.neuroimage.2010.07.034.

[51] McIntosh AR, Lobaugh NJ. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 2004;23:S250–S263.

[52] Escofier-Cordier B. L'Analyse des Correspondences. Unveröffentlichte Dissertation. Université de Rennes, 1965.

[53] Benzécri JP, Robert J. Histoire et préhistoire de l'analyse des données. Partie I-La Préhistoire. Cah Anal Données 1976;1:9–32.

[54] Greenacre MJ. Theory and Applications of Correspondence Analysis. Academic Press; 1984.

[55] Greenacre M. Correspondence analysis in practice. CRC press; 2017.

[56] Abdi H. Discriminant correspondence analysis. Encycl. Meas. Stat., 2007, p. 284–90.

[57] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

[58] Beaton D, Fatt CRC, Abdi H. An ExPosition of multivariate analysis with the singular value decomposition in R. Comput Stat Data Anal 2014;72:176–89. doi:http://dx.doi.org/10.1016/j.csda.2013.11.006.

[59] Beaton D, Rieck J, Abdi H. TInPosition: Inference Tests for TExPosition. 2013.

[60] Beaton D, Kriegsman M, ADNI, Dunlop JP, Filbey FM, Abdi H. Partial Least Squares for mixed-data types: An application for imaging genetics. Mult. Facets Partial Least Sq. Methods, vol. H. Abdi, V. Esposito Vinzi, G. Russolillo, S. Saporta, L. Trinchera, (Eds.), New York, NY, USA: Springer Verlag; 2016, p. 63–78.

[61] Habib R, McIntosh AR, Wheeler MA, Tulving E. Memory encoding and hippocampally-based novelty/familiarity discrimination networks. Neuropsychologia 2003;41:271–9. doi:10.1016/S0028-3932(02)00160-4.

[62] Nair HP, Berndt JD, Barrett D, Gonzalez-Lima F. Maturation of Extinction Behavior in Infant Rats: Large-Scale Regional Interactions with Medial Prefrontal Cortex, Orbitofrontal Cortex, and Anterior Cingulate Cortex. J Neurosci 2001;21:4400–7.

[63] McIntosh AR, Bookstein FL, Haxby JV, Grady CL. Spatial Pattern Analysis of Functional Brain Images Using Partial Least Squares. NeuroImage 1996;3:143–57. doi:10.1006/nimg.1996.0016.

[64] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 2007;81:559–75.

[65] Melville S. NCBI2R: Queries and Annotates SNPs, Genes and Pathway from NCBI. 2015.

[66] Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 2009;4:1184–91. doi:10.1038/nprot.2009.97.

[67] Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinforma Oxf Engl 2005;21:3439–40. doi:10.1093/bioinformatics/bti525.

[68] Efron B. Bootstrap Methods: Another Look at the Jackknife. Ann Stat 1979;7:1–26.

[69] Hao X, Li C, Du L, Yao X, Yan J, Risacher SL, et al. Mining Outcome-relevant Brain Imaging Genetic Associations via Three-way Sparse Canonical Correlation Analysis in Alzheimer's Disease. Sci Rep 2017;7:srep44272. doi:10.1038/srep44272.

[70] Edelaar P, Roques S, Hobson EA, Gonçalves da Silva A, Avery ML, Russello MA, et al. Shared genetic diversity across the global invasive range of the monk parakeet suggests a common restricted geographic origin and the possibility of convergent selection. Mol Ecol 2015;24:2164–76. doi:10.1111/mec.13157.

[71] Greenacre MJ, Degos L. Correspondence analysis of HLA gene frequency data from 124 population samples. Am J Hum Genet 1977;29:60–75.

[72] Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, et al. A Back Migration from Asia to Sub-Saharan Africa Is Supported by High-Resolution Analysis of Human Y-Chromosome Haplotypes. Am J Hum Genet 2002;70:1197–214. doi:10.1086/340257.

[73] Suchanek R, Owczarek A, Kowalczyk M, Kucia K, Kowalski J. Association Between C-281A and Val66met Functional Polymorphisms of BDNF Gene and Risk of Recurrent Major Depressive Disorder in Polish Population. J Mol Neurosci 2011;43:524–30. doi:10.1007/s12031-010-9478-y.

[74] Paul-Samojedny M, Kowalczyk M, Suchanek R, Owczarek A, Fila-Danilow A, Szczygiel A, et al. Functional Polymorphism in the Interleukin-6 and Interleukin-10 Genes in Patients with Paranoid Schizophrenia —A Case-Control Study. J Mol Neurosci 2010;42:112–9. doi:10.1007/s12031-010-9365-6.

[75] Onori N, Turchi C, Solito G, Gesuita R, Buscemi L, Tagliabracci A. GABRA2 and Alcohol Use Disorders: No Evidence of an Association in an Italian Case–Control Study. Alcohol Clin Exp Res 2010;34:659–668. doi:10.1111/j.1530-0277.2009.01135.x.

[76] Linnertz C, Anderson L, Gottschalk W, Crenshaw D, Lutz MW, Allen J, et al. The cis-regulatory effect of an Alzheimer's disease-associated poly-T locus on expression of TOMM40 and apolipoprotein E genes. Alzheimers Dement 2014;10:541–51. doi:10.1016/j.jalz.2013.08.280.

[77] Ferrari R, Hernandez DG, Nalls MA, Rohrer JD, Ramasamy A, Kwok JBJ, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neurol 2014;13:686–99. doi:10.1016/S1474-4422(14)70065-1.

[78] Chen JA, Wang Q, Davis-Turak J, Li Y, Karydas AM, Hsu SC, et al. A Multiancestral Genome-Wide Exome Array Study of Alzheimer Disease, Frontotemporal Dementia, and Progressive Supranuclear Palsy. JAMA Neurol 2015.

[79] Udayar V, Buggia-Prévot V, Guerreiro RL, Siegel G, Rambabu N, Soohoo AL, et al. A Paired RNAi and RabGAP Overexpression Screen Identifies Rab11 as a Regulator of β-Amyloid Production. Cell Rep 2013;5:1536–51. doi:10.1016/j.celrep.2013.12.005.

[80] Bertram L, Tanzi RE. Genome-wide association studies in Alzheimer's disease. Hum Mol Genet 2009;18:R137–45. doi:10.1093/hmg/ddp406.

[81] Cruchaga C, Chakraverty S, Mayo K, Vallania FLM, Mitra RD, Faber K, et al. Rare Variants in APP, PSEN1 and PSEN2 Increase Risk for AD in Late-Onset Alzheimer's Disease Families. PLoS ONE 2012;7:e31039. doi:10.1371/journal.pone.0031039.

[82] Scheuner D, Eckman C, Jensen M, Song X, Citron M, Suzuki N, et al. Secreted amyloid β–protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. Nat Med 1996;2:864–70. doi:10.1038/nm0896-864.

[83] Zhao J, Zhu Y, Yang J, Li L, Wu H, De Jager PL, et al. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. Alzheimers Dement 2017;13:674–88. doi:10.1016/j.jalz.2016.10.004.

[84] Vounou M, Janousova E, Wolz R, Stein JL, Thompson PM, Rueckert D, et al. Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease. NeuroImage 2012;60:700–16. doi:10.1016/j.neuroimage.2011.12.029.

[85] Huang C, Thompson P, Wang Y, Yu Y, Zhang J, Kong D, et al. FGWAS: Functional genome wide association analysis. NeuroImage 2017;159:107–21. doi:10.1016/j.neuroimage.2017.07.030.

[86] Yang T, Wang J, Sun Q, Hibar DP, Jahanshad N, Liu L, et al. Detecting genetic risk factors for Alzheimer's disease in whole genome sequence data via Lasso screening. 2015 IEEE 12th Int. Symp. Biomed. Imaging ISBI, 2015, p. 985–9. doi:10.1109/ISBI.2015.7164036.

[87] Sherva R, Tripodis Y, Bennett DA, Chibnik LB, Crane PK, de Jager PL, et al. Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. Alzheimers Dement 2014;10:45–52. doi:10.1016/j.jalz.2013.01.008.

[88] Gillingwater TH, Wishart TM. Mechanisms underlying synaptic vulnerability and degeneration in neurodegenerative disease. Neuropathol Appl Neurobiol 2013;39:320–34. doi:10.1111/nan.12014.

[89] French L, Ma T, Oh H, Tseng GC, Sibille E. Age-Related Gene Expression in the Frontal Cortex Suggests Synaptic Function Changes in Specific Inhibitory Neuron Subtypes. Front Aging Neurosci 2017;9. doi:10.3389/fnagi.2017.00162.

[90] Parcerisas A, Rubio SE, Muhaisen A, Gómez-Ramos A, Pujadas L, Puiggros M, et al. Somatic Signature of Brain-Specific Single Nucleotide Variations in Sporadic Alzheimer's Disease. J Alzheimers Dis 2014;42:1357–82. doi:10.3233/JAD-140891.

[91] Zhang L, Yang S, Wennmann DO, Chen Y, Kremerskothen J, Dong J. KIBRA: In the brain and beyond. Cell Signal 2014;26:1392–9. doi:10.1016/j.cellsig.2014.02.023.

[92] Witte AV, Köbe T, Kerti L, Rujescu D, Flöel A. Impact of KIBRA Polymorphism on Memory Function and the Hippocampus in Older Adults. Neuropsychopharmacology 2015;41:npp2015203. doi:10.1038/npp.2015.203.

[93] Lee J-M, Gillis T, Mysore JS, Ramos EM, Myers RH, Hayden MR, et al. Common SNP-Based Haplotype Analysis of the 4p16.3 Huntington Disease Gene Region. Am J Hum Genet 2012;90:434–44. doi:10.1016/j.ajhg.2012.01.005.

[94] Ganief T, Gqamana P, Garnett S, Hoare J, Stein DJ, Joska J, et al. Quantitative proteomic analysis of HIV-1 Tat-induced dysregulation in SH-SY5Y neuroblastoma cells. PROTEOMICS 2017;17:n/a-n/a. doi:10.1002/pmic.201600236.

[95] Zhang M, Mu H, Shang Z, Kang K, Lv H, Duan L, et al. Genome-wide pathway-based association analysis identifies risk pathways associated with Parkinson's disease. Neuroscience 2017;340:398–410. doi:10.1016/j.neuroscience.2016.11.004.

[96] Durrenberger PF, Fernando FS, Kashefi SN, Bonnert TP, Seilhean D, Nait-Oumesmar B, et al. Common mechanisms in neurodegeneration and neuroinflammation: a BrainNet Europe gene expression microarray study. J Neural Transm 2015;122:1055–68. doi:10.1007/s00702-014-1293-0.

[97] Spellman DS, Wildsmith KR, Honigberg LA, Tuefferd M, Baker D, Raghavan N, et al. Development and evaluation of a multiplexed mass spectrometry based assay for measuring candidate peptide biomarkers in Alzheimer's Disease Neuroimaging Initiative (ADNI) CSF. PROTEOMICS – Clin Appl 2015;9:715–31. doi:10.1002/prca.201400178.

[98] Esteras N, Bartolomé F, Alquézar C, Antequera D, Muñoz Ú, Carro E, et al. Altered cell cycle-related gene expression in brain and lymphocytes from a transgenic mouse model of Alzheimer's disease [amyloid precursor protein/presenilin 1 (PS1)]. Eur J Neurosci 2012;36:2609–18. doi:10.1111/j.1460-9568.2012.08178.x.

[99] Yokota T, Mishra M, Akatsu H, Tani Y, Miyauchi T, Yamamoto T, et al. Brain site-specific gene expression analysis in Alzheimer's disease patients. Eur J Clin Invest 2006;36:820–30. doi:10.1111/j.1365-2362.2006.01722.x.

[100] Niwa M, Cash-Padgett T, Kubo K-I, Saito A, Ishii K, Sumitomo A, et al. DISC1 a key molecular lead in psychiatry and neurodevelopment: No-More Disrupted-in-Schizophrenia. Mol Psychiatry 2016;21:1488–9. doi:10.1038/mp.2016.154.

[101] Beecham GW, Martin ER, Li Y-J, Slifer MA, Gilbert JR, Haines JL, et al. Genome-wide Association Study Implicates a Chromosome 12 Risk Locus for Late-Onset Alzheimer Disease. Am J Hum Genet 2009;84:35–43. doi:10.1016/j.ajhg.2008.12.008.

[102] Shahani N, Seshadri S, Jaaro-Peled H, Ishizuka K, Hirota-Tsuyada Y, Wang Q, et al. DISC1 regulates trafficking and processing of APP and Aβ generation. Mol Psychiatry 2014. doi:10.1038/mp.2014.100.

[103] Carless MA, Glahn DC, Johnson MP, Curran JE, Bozaoglu K, Dyer TD, et al. Impact of DISC1 variation on neuroanatomical and neurocognitive phenotypes. Mol Psychiatry 2011;16:1096–1063. doi:10.1038/mp.2011.37.

[104] Consortium TAG. Seven new loci associated with age-related macular degeneration. Nat Genet 2013;45:433–9. doi:10.1038/ng.2578.

[105] Logue MW, Schu M, Vardarajan BN, Farrell J, Lunetta KL, Jun G, et al. A search for age-related macular degeneration risk variants in Alzheimer disease genes and pathways. Neurobiol Aging 2014;35:1510.e7-1510.e18. doi:10.1016/j.neurobiolaging.2013.12.007.

Supplemental Material

Table S1
Dx DiCA Significant genotypes

(a) Dx DiCA Component 1 (AD vs. MCI) – significant genotypes

rsid	genotype	BSR –1	BSR –2	Chr	Gene Symbol
rs1061622	GG	-5.32	-1.059	1	*TNFRSF1B*
rs6696924	GG	-5.516	-0.177	1	*C8A*
rs679350	TT	-5.007	-0.244	1	*C8A*
rs1522551	TT	-5.075	-0.197	3	
rs7626449	AA	-5.332	0.454	3	
rs13157174	GG	-5.17	-0.074	5	*LINC02216*
rs2875382	TT	-5.057	-1.29	6	*MRAP2*
rs4501410	GG	-6.05	1.185	6	
rs11783013	GG	-5.498	-0.397	8	
rs902466	TT	-5.44	0.23	10	*ARHGAP19-SLIT1*
rs16977252	GG	-5.617	-0.471	15	*AKAP13*
rs12934725	GG	-5.23	0.536	16	*RBFOX1*
rs9952815	GG	-6.027	0.745	18	*NOL4*
rs283168	TT	-5.331	-0.389	19	
rs916326	GG	5.356	0.315	20	*PTPRT*
rs13054435	AA	-5.541	-0.213	22	*NUP50-AS1*
rs4820946	CC	-5.641	-0.5	22	*MIR3928*

(b) Dx DiCA Component 2 (CON vs. {AD & MCI}) – significant genotypes

rsid	genotype	BSR - 1	BSR - 2	Chr	Gene
rs1115763	CC	-1.001	5.97	2	*AC007682.1*
rs12618595	AC+AA	1.418	5.474	2	*OSBPL6*
rs1470524	TT	-1.144	5.178	2	
rs1387089	CT+CC	-0.701	5.015	3	
rs1387094	GT+GG	-0.541	5.004	3	
rs4685465	GT+GG	-0.765	5.136	3	
rs9821034	GG	0.905	5.37	3	
rs3846336	AG+AA	0.462	5.329	4	*CCDC149*
rs6448119	TT	0.027	5.711	4	
rs6812046	TT	0.355	5.462	4	
rs6851636	GG	-0.371	5.457	4	
rs7656406	AA	0.184	5.48	4	
rs6882277	GG	0.23	5.867	5	
rs13236754	GG	-0.163	5.234	7	*ZNF398*
rs1557664	GG	1.43	5.924	7	
rs9640538	AA	-0.297	5.487	7	
rs10814567	GG	0.16	5.36	9	*POLR1E*
rs10814571	AA	-0.128	6.816	9	*POLR1E*
rs1590255	CA+CC	-0.732	5.042	9	
rs3739574	TT	-0.121	6.805	9	*POLR1E*
rs7468695	GA+GG	-1.168	5.073	9	
rs11598825	AA	-0.487	5.257	10	*ITIH5*
rs947696	TT	1.4	5.473	10	
rs11160481	CC	0.013	5.017	14	
rs1243473	TT	0.439	5.782	14	*ARHGEF40*
rs4635275	AA	0.002	5.028	14	
rs7146951	GG	0.005	5.764	14	
rs9806693	AA	-0.668	5.969	15	*MORF4L1*
rs4789240	TT	1.051	5.089	17	*SDK2*
rs9892996	AA	-1.531	6.928	17	
rs9949152	CC	-3.019	5.005	18	
rs2075650	AA	3.295	-5.584	19	*TOMM40*
rs7251241	AA	0.937	5.282	19	*UNC13A*
rs2830052	CC	-0.086	5.614	21	*APP*
rs8141950	TC+TT	0.767	5.674	22	*PARVB*

Note. Significant genotypes from Dx DiCA Components 1 (a) and 2 (b). Gene symbols via `NCBI2R`. Component 1 separated AD from MCI, thus, most genotypes here are more associated

with AD than MCI; the exception is rs916326 where GG is more associated with MCI than AD. Component 2 separates CON from {AD & MCI}, thus, most genotypes here are more associated with disorder status (AD or MCI) than CON; the exception is rs2075650 where AA is more associated with CON than disease status.

Table S2
APOE DiCA Significant genotypes

(a) *APOE* DiCA Component 1 (Presence vs. Absence of E4) – significant genotypes

rsid	genotype	BSR - 1	BSR - 2	Chr	Gene Symbol
rs10830213	GG	6.537	-0.724	11	*RAB38*
rs157580	AA	-7.842	-2.918	19	*TOMM40*
rs157580	GG	14.612	-2.13	19	*TOMM40*
rs2075650	AA	15.323	-0.619	19	*TOMM40*
rs405509	AA	-6.101	-2.567	19	*APOE*
rs405509	CC	6.701	-1.244	19	*APOE*
rs439401	CC	-8.703	-2.63	19	
rs6859	AA	-5.614	-3.096	19	*NECTIN2/PVRL2*
rs8106922	AA	-7.514	-3.49	19	*TOMM40*
rs8106922	GG	20.928	-4.43	19	*TOMM40*

(b) *APOE* DiCA Component 2 (2 E4 alleles vs. {0 or 1}) – significant genotypes

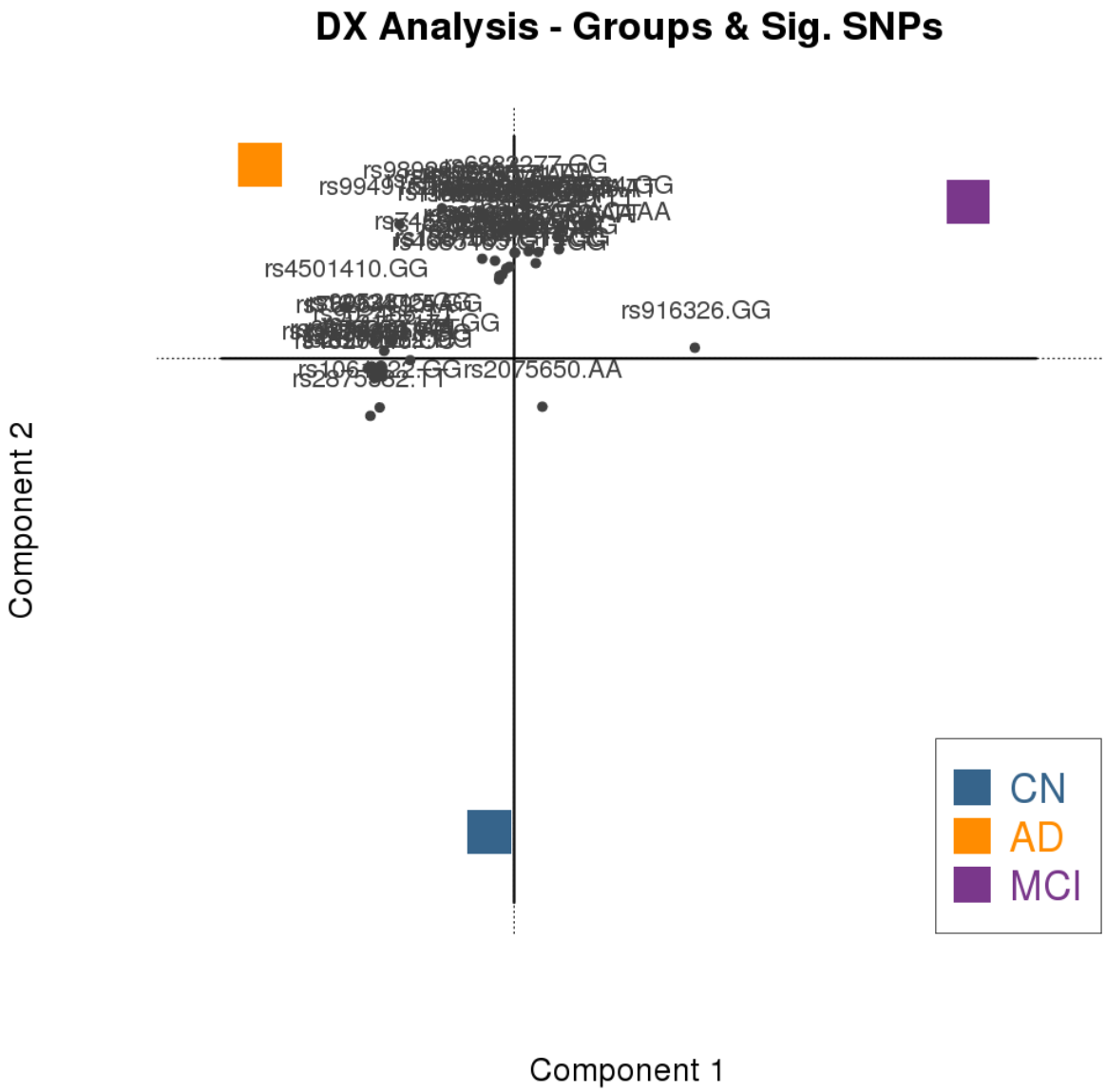
rsid	genotype	BSR - 1	BSR - 2	Chr	Gene Symbol
rs2000072	AA	-0.211	5.066	1	*LINC00624*
rs4253890	CC	-1.077	5.985	1	*PTPN14*
rs6681032	TC+TT	0.157	5.439	1	
rs6703696	GA+GG	0.169	5.604	1	
rs7541019	GG	0.576	5.069	1	*TSNAX-DISC1,DISC1*
rs11899115	AA	-0.64	5.996	2	
rs13009482	CT+CC	-1.383	5.696	2	
rs722963	TT	-0.288	5.093	2	
rs10019637	CC	-0.964	5.412	4	
rs10804966	AA	-0.151	5.267	4	*EVC*
rs300574	TT	0.275	5.86	4	*SPRY1*
rs7678082	TC+TT	-0.105	5.196	4	*WWC2*
rs7681283	GG	-0.168	5.278	4	*EVC*
rs868082	TC+TT	-0.187	6.13	4	
rs10041935	CC	-0.895	5.75	5	
rs236444	TT	-1.938	7.117	6	*CPNE5*
rs1673206	TT	-1.294	5.026	7	
rs6960851	AA	-0.713	5.079	7	
rs537941	TT	-0.73	5.27	8	*NCALD*
rs1492598	AA	-0.933	5.119	9	
rs4935847	TT	-0.247	6.085	11	
rs1647147	GG	-1.396	5.407	12	
rs16928445	TT	0.609	5.015	12	*LRMP*
rs4759955	TT	-0.49	5.903	12	*TMEM132D*
rs6582412	AA	-1.433	5.113	12	
rs944838	CC	0.373	5.917	13	*GPC6*
rs9549831	AA	0.454	5.082	13	
rs4905290	GG	-0.457	5.019	14	*CLMN*
rs6573852	GG	0.438	5.294	14	*RAD51B*
rs7148010	TT	1.018	5.112	14	*SMOC1*
rs10519492	GA+GG	-0.468	5.537	15	
rs6496431	GG	-0.087	5.072	15	
rs714900	TC+TT	-0.627	5.611	15	
rs7177541	AA	-0.664	5.299	15	
rs4500815	AA	-0.28	5.3	18	*CTIF*
rs9955327	CC	-1.958	5.021	18	*CELF4*
rs10423685	TT	-0.479	5.053	19	*ZNF600*
rs157580	GA	4.603	8.591	19	*TOMM40*
rs6509238	CC	0.056	5.038	19	

rs8106922	GA	3.703	9.456	19	*TOMM40*

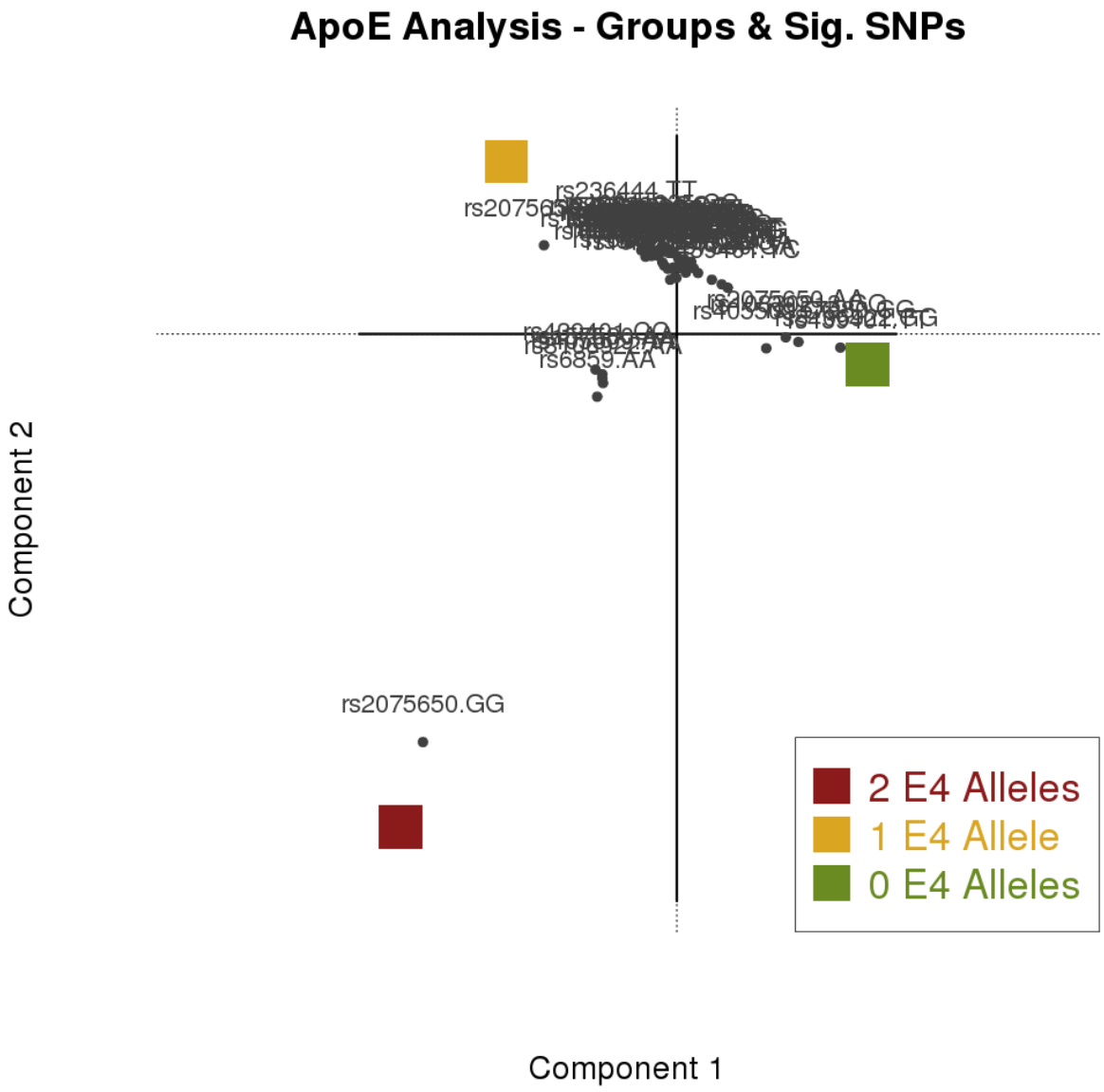
(c) *APOE* DiCA Significant genotypes on both Components.

rsid	genotype	BSR - 1	BSR - 2	Chr
rs2075650	GA	-17.593	7.374	19
rs2075650	GG	-25.132	-13.27	19
rs439401	TC	5.052	7.92	19
rs439401	TT	25.848	-5.79	19

Note. Significant genotypes from *APOE* DiCA Components 1 (a), 2 (b), and on both Components (c). Gene symbols via `NCBI2R`. Component 1 separated presence from absence of E4 alleles. Genotypes that have a negative bootstrap ratio (BSR) were more associated with the presence of an E4 allele. Component 2 separated, essentially, the 2 E4 alleles from the other (0 or 1) E4 alleles. In (b) all genotypes were more related to the *absence* of 2 E4 alleles. In (c) these genotypes contribute to both components and suggest that these genotypes are in very high linkage disequilibrium with *APOE* (note that the GG genotype of rs2075650 strongly contributes to both components and in the same direction as the *presence* 2 E4 alleles).



Supplemental Figure 1. The component map from the Dx-DICA (discovery phase). The component map shows significant genotypes and the group configuration to illustrate—as a biplot—the relationship between the genotypes and groups.



Supplemental Figure 2. The component map from the *APOE*-DICA (discovery phase). The component map shows significant genotypes and the group configuration to illustrate—as a biplot—the relationship between the genotypes and groups.