Title
Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease

Permalink
https://escholarship.org/uc/item/2xh0j2p6

Journal
Circulation, 131(23)

ISSN
0009-7322

Authors
Desikan, RS
Schork, AJ
Wang, Y
et al.

Publication Date
2015

DOI
10.1161/CIRCULATIONAHA.115.015489

Peer reviewed
Late-onset Alzheimer disease (AD) is the most common form of dementia, with an estimated prevalence of 30 million people worldwide, a number that is expected to quadruple in the next 40 years.\(^1\) Given the absence of disease-modifying therapies and increasing awareness that symptoms develop over many years, there is significant interest in identifying effective strategies for AD prevention. Delaying dementia onset by a modest 2 years could potentially lower the world-wide prevalence of AD by >22 million cases over the next 40 years, resulting in significant societal savings.\(^1\)

### Clinical Perspective on p 2069

A growing body of evidence suggests an association between AD and potentially modifiable processes, including dyslipidemia and inflammation. In observational studies, high serum cholesterol levels have been associated with increased risk of AD.\(^2,3\) and molecular\(^4\) and biomarker findings\(^5\) suggest that phospholipids may play an integral role in modulating AD-associated pathogenesis. Complement factors and activated microglia are established histopathological features in brains of AD patients,\(^6\) and epidemiological studies in older individuals indicate that high serum levels of inflammatory proteins are associated with cognitive decline\(^7\) and may predict dementia risk.\(^8\) Genome-wide association studies (GWASs) in late-onset AD have replicated the established association with apolipoprotein E (APOE) and identified single-nucleotide polymorphisms (SNPs) implicated in lipid metabolism such as \(CLU\) and \(ABCA7\) and inflammatory processes such as \(CR1\) and \(HLA-DRB5\).\(^9,10\) In addition, a rare sequence variant in \(TREM-2\) with known anti-inflammatory function has recently been identified as conferring increased risk for AD.\(^11,12\) Taken together, these findings suggest that processes involved with lipid metabolism and inflammation may also affect AD pathogenesis.

Combining GWASs from multiple disorders and phenotypes provides insights into genetic pleiotropy (defined as a single gene or variant being associated with >1 distinct phenotype) and could elucidate shared pathobiology. Using this approach, we have recently reported genetic overlap between a number of diseases and phenotypes and identified novel common variants associated with schizophrenia,\(^13,14\) bipolar disorder,\(^13\) prostate cancer,\(^15\) hypertension,\(^16\) and primary sclerosing cholangitis.\(^17\) Here, we applied this method to AD, taking advantage of several large GWASs,\(^18-20\) to identify SNPs associating with clinically diagnosed AD, C-reactive protein (CRP) levels, and plasma lipid levels (specifically triglycerides, high-density lipoprotein [HDL], and low-density lipoproteins [LDL]).

### Methods

#### Participant Samples

We evaluated complete GWAS results in the form of summary statistics (\(P\) values and odds ratios [ORs]) for clinically diagnosed AD,\(^18\) CRP levels,\(^19\) and plasma lipid levels (triglycerides, HDL, and LDL); Table 1). The CRP GWAS summary statistic data consisted of 82725 individuals drawn from 25 studies with genotyped or imputed data at 2671742 SNPs (for additional details, see Reference 19). The plasma lipid GWAS summary statistic data consisted of 188577 individuals with genotyped or imputed data at 2508375 SNPs (for additional details, see reference 20). We obtained publicly available AD GWAS summary statistic data from the International Genomics of Alzheimer’s Disease Project (IGAP stage 1 and 2; for additional details, see the online-only Data Supplement and Reference 18). We used IGAP stage 1 as our discovery cohort, which consisted of 17008 AD cases (mean age, 74.7±7.7 years; 59.4% female) and 37154 controls (mean age, 76.3±8.1 years; 58.6% female) drawn from 4 different consortia across North America and Europe with genotyped or imputed data at 7055811 SNPs (for a description of the AD cases and controls within the IGAP stage 1 substudies, see Reference 18).

To confirm our findings from IGAP stage 1, we assessed the \(P\) values of pleiotropic SNPs (conditional false discovery rate [FDR] <0.05; see Statistical Analysis) from the discovery analyses in 3 independent
Table 1. Summary Data From All GWASs Used in the Present Study

<table>
<thead>
<tr>
<th>Disease/Trait</th>
<th>n</th>
<th>SNPs, n</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD, IGAP stages 1+2</td>
<td>74,046</td>
<td>7,055,881 (stage 1)+11,632 (stage 2)</td>
<td>Lambert et al14</td>
</tr>
<tr>
<td>AD, deCODE</td>
<td>6,717</td>
<td>Whole-genome sequencing</td>
<td>Jonsson et al12</td>
</tr>
<tr>
<td>AD, DemGene</td>
<td>2,015</td>
<td>6,933,777</td>
<td>N/A</td>
</tr>
<tr>
<td>TG</td>
<td>1,885,577</td>
<td>2,508,369</td>
<td>Teslovich et al13</td>
</tr>
<tr>
<td>LDL</td>
<td>1,885,577</td>
<td>2,508,375</td>
<td></td>
</tr>
<tr>
<td>HDL</td>
<td>1,885,577</td>
<td>2,508,370</td>
<td></td>
</tr>
<tr>
<td>CRP</td>
<td>8,272</td>
<td>2,671,742</td>
<td>Dehghan et al19</td>
</tr>
</tbody>
</table>

AD indicates Alzheimer disease; CRP, C-reactive protein; HDL, high-density lipoprotein; IGAP, International Genomics of Alzheimer’s Disease Project; LDL, low-density lipoprotein; and TG, triglycerides.

Statistical Analysis

Using recently developed statistical methods to evaluate pleiotropic effects,13–17 we evaluated SNPs associating with AD (discovery cohort, IGAP stage 1) and CRP levels, as well as AD and plasma lipid levels. For given associated phenotypes A and B, pleiotropic enrichment of phenotype A with phenotype B exists if the proportion of SNPs or genes associated with phenotype A increases as a function of increased association with phenotype B. To assess for enrichment, we constructed fold-enrichment plots of nominal –log10(P) values for all AD SNPs and for subsets of SNPs determined by the significance of their associations with CRP and plasma lipids. We also used conditional quantile-quantile plots, which are complementary to fold-enrichment plots and provide visualization of polygenic enrichment (for additional details, see the online-only Data Supplement). In fold-enrichment plots, the presence of enrichment is reflected as an upward deflection of the curve for phenotype A if the degree of deflection from the expected null line is dependent on the degree of association with phenotype B. To assess for polygenic effects below the standard GWAS significance threshold, we focused the fold-enrichment plots on SNPs with nominal –log10(P)<7.3 (corresponding to P>5×10−6).

The enrichment seen can be directly interpreted in terms of true discovery rate (true discovery rate=1−FDR; for additional details, see the online-only Data Supplement).

To identify specific loci, we computed conditional FDRs.13,14 The standard FDR framework derives from a model that assumes that the distribution of test statistics in a GWAS can be formulated as a mixture of null and nonnull effects, with true associations (nonnull effects) having more extreme test statistics, on average, than false associations (null effects). The FDR can be interpreted as the probability that an SNP is null given that its P value is as small or smaller than its observed P value. The conditional FDR is an extension of the standard FDR, which incorporates information from GWAS summary statistics of a second phenotype to adjust its significance level. The conditional FDR is defined as the probability that an SNP is null in the first phenotype given that the P values in the first and second phenotypes are as small as or smaller than the observed values. It is important to note that ranking SNPs by standard FDRs or by P values gives the same ordering of SNPs. In contrast, if the primary and secondary phenotypes are related genetically, conditional FDR reorders SNPs and results in a different ranking than that based on P values alone. We used an overall FDR threshold of <0.05, which means 5 expected false discoveries per 100 reported. Additionally, we constructed Manhattan plots based on the ranking of conditional FDR to illustrate the genomic location. In all analyses, we controlled for the effects of genomic inflation by using intergenic SNPs (see the online-only Data Supplement). Detailed information on fold-enrichment and conditional quantile-quantile plots, Manhattan plots, and conditional FDR can be found in the online-only Data Supplement and prior reports.13–17

For loci with conditional FDR <0.05, we performed a fixed-effects, inverse variance-weighted meta-analysis18 across all available AD cohorts (IGAP stages 1 and 2, deCODE, and DemGene; total: n=29,054 AD cases and 114,824 healthy controls) using the R package meta (http://CRAN.R-project.org/package=meta). Briefly, the fixed-effects, inverse variance-weighted meta-analysis summarizes the combined statistical support across independent studies under the assumption of homogeneity of effects. Individual study β estimates (log ORs) are averaged, weighted by the estimated standard error.20 The IGAP stage 1 and 2 β estimates and standard errors were obtained from the publicly available summary statistics (for additional details, see the online-only Data Supplement and the Supplementary Note in.
Reference18). For the DeCODE and DemGene cohorts, β estimates and standard errors were estimated via logistic regression, predicting AD case/control status from SNP risk alleles count and adjusting for appropriate covariates, including principal components.

For the gene expression analyses, we focused on transcript expression (total RNA levels) of genes closest (within 500 kb) to the SNPs reaching genome-wide significance in our meta-analysis. Using logistic regression, we examined whether transcript expression of these genes significantly differed between AD cases and controls.

Results

We observed SNP enrichment for AD (IGAP stage 1, discovery cohort) across different levels of significance, with CRP, triglycerides, HDL, and LDL levels indicating a genetic association between AD and the 4 cardiovascular phenotypes (Figure 1). For progressively stringent P value thresholds for AD SNPs [ie, increasing values of nominal \(-\log_{10}(P)\)], we found at least 50-fold enrichment using CRP, 30-fold enrichment using triglycerides, 20-fold enrichment using HDL, and 40-fold enrichment using LDL (Figure 1). Conditional quantile-quantile plots similarly demonstrated polygenic enrichment in AD as a function of CRP and plasma lipids (Figure I in the online-only Data Supplement).

To identify AD-associated polymorphisms that are more likely to replicate, we ranked IGAP stage 1 AD SNPs conditional on their genetic association with CRP and plasma lipids (conditional FDR). We restricted our analyses to SNPs found in both IGAP stages 1 and 2 and focused on those AD variants that have not previously been described at a genome-wide significant level. At a conditional FDR < 0.05, we found 55 AD susceptibility loci from IGAP stage 1 (Figure 2 and Table I in the online-only Data Supplement). For these 55 loci, we performed a meta-analysis across all available AD cohorts and found 2 novel genome-wide significant \((P<5\times 10^{-8})\) loci associated with increased risk for AD (Table 2). These 2 variants are rs13113697 (chromosome 4; closest gene, HS3ST1; conditioning trait, triglycerides; reference allele, T; \(OR=1.07\); 95% CI=1.05–1.11; \(P=2.86\times 10^{-8}\); Figures 3A and 4A) and rs7920721 (chromosome 10; closest gene, ECHDC3; conditioning trait, triglycerides; reference allele, G; \(OR=1.07\); 95% CI=1.04–1.11; \(P=3.38\times 10^{-8}\); Figures 3B and 4B).

The meta-analysis also revealed 3 suggestive AD susceptibility loci with values of \(P<1\times 10^{-6}\) (Table 3 and Figure II in the online-only Data Supplement). These 3 loci are rs7396366 (on chromosome 11; closest gene, AP2A2; conditioning trait, CRP; reference allele, C; \(OR=0.94\); 95% CI=0.92–0.96;...
We additionally evaluated the directionality of allelic effects in SNPs associated with AD and the 4 cardiovascular phenotypes (SNPs with conditional FDR <0.05). Across all 55 shared loci, we found the same direction of effect between SNPs associated with AD and CRP in 72% (18 of 25; \(P=0.02\)), HDL in 40% (4 of 10; \(P=0.62\)), LDL in 20% (1 of 5; \(P=0.81\)), and triglycerides in 40% (6 of 15; \(P=0.69\); Table I in the online-only Data Supplement). For \(HS3ST1\) and \(ECHD3\) variants, we found an opposite direction of allelic effect between increased AD risk and triglyceride levels (Table I in the online-only Data Supplement).

We assessed whether \(HS3ST1\) and \(ECHDC3\) transcript levels are altered in AD brains compared with control brains (Gene Expression Omnibus data set GSE 4472). We found significantly decreased \(HS3ST1\) transcript expression (standardized \(\beta\) coefficient=-0.09201; SE=0.01864; \(P=9.99\times10^{-7}\)) and significantly increased \(ECHDC3\) transcript expression (standardized \(\beta\) coefficient=0.12715; SE=0.01829; \(P=8.32\times10^{-12}\)) in AD brains compared with control brains.

### Table 2. New Loci Reaching Genome-Wide Significance at Conditional FDR <0.05 (ORs Provided for the Reference Allele)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Nearest Gene</th>
<th>Reference Allele</th>
<th>Associated Phenotype</th>
<th>Minimum Conditional FDR</th>
<th>IGAP Stage 1+2 OR (95%CI)</th>
<th>IGAP Stage 1+2 P Value</th>
<th>deCODE OR (95%CI)</th>
<th>deCODE P Value</th>
<th>DemGene OR (95%CI)</th>
<th>DemGene P Value</th>
<th>Meta-Analysis OR (95%CI)</th>
<th>Meta-Analysis P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs13113697</td>
<td>4</td>
<td>(HS3ST1)</td>
<td>(TG)</td>
<td></td>
<td>9.56×10⁻³</td>
<td>5.03×10⁻⁷ (1.04–1.10)</td>
<td>1.07</td>
<td>0.031</td>
<td>1.07</td>
<td>0.088</td>
<td>1.13</td>
<td>2.86×10⁻⁸ (1.05–1.11)</td>
<td>1.07</td>
</tr>
<tr>
<td>rs7920721</td>
<td>10</td>
<td>(ECHDC3)</td>
<td>(TG)</td>
<td></td>
<td>4.09×10⁻²</td>
<td>2.89×10⁻⁷ (0.99–1.11)</td>
<td>1.07</td>
<td>0.12</td>
<td>1.05</td>
<td>0.08</td>
<td>1.12</td>
<td>3.38×10⁻⁸ (1.04–1.11)</td>
<td>1.07</td>
</tr>
</tbody>
</table>

\(\text{CI}\) indicates confidence interval; FDR, false discovery rate; IGAP, International Genomics of Alzheimer’s Disease Project; OR, odds ratio; SNP, single-nucleotide polymorphism; and \(TG\), triglycerides.

**Figure 2.** Conditional Manhattan plot of conditional \(-\log_{10}(FDR)\) values for Alzheimer disease (AD) alone (International Genomics of Alzheimer’s Disease Project [IGAP] stage 1 AD cohort; black) and AD given C-reactive protein (CRP; AD-CRP; green), triglycerides (TG; AD-TG; aquamarine), high-density lipoprotein (HDL; AD-HDL; orange), and low-density lipoprotein (LDL; AD-LDL; red). Single-nucleotide polymorphisms (SNPs) with conditional \(-\log_{10}(FDR)\) >1.3 (ie, FDR <0.05) are shown with large points. A black line around the large points indicates the most significant SNP in each linkage disequilibrium block, and this SNP was annotated with the closest gene, which is listed above the symbols in each locus. For additional details, see the online-only Data Supplement.

**Discussion**

In this study, we show that polymorphisms associated with CRP and plasma lipids (triglycerides, HDL, and LDL) are also associated with increased risk for AD (genetic pleiotropy). We found that genetic enrichment in AD based on SNP association with cardiovascular phenotypes results in improved statistical power for gene discovery. By conditioning on polymorphisms associated with CRP and plasma lipid levels, we identified 55 AD susceptibility loci. In meta-analyses across 4 independent cohorts, we found that 2 of these risk variants, namely rs13113697 (on chromosome 4; closest gene, \(HS3ST1\)) and rs7920721 (on chromosome 10; closest gene, \(ECHDC3\)), were genome-wide significant. We additionally observed that \(HS3ST1\) and \(ECHDC3\) transcript expression was different in AD brains compared with control brains.

Our findings provide novel insights into the relationship among AD pathogenesis, inflammation, and dyslipidemia beyond the known loci associated with AD. We found a consistent direction of allelic effect between SNPs associated with AD risk and CRP levels, indicating overlapping pathobiology...
between AD and inflammation. These results are consistent
with the hypothesis that inflammatory mechanisms influence
AD pathogenesis and may have implications for treatment and
prevention strategies in AD. On the other hand, we did not
find a consistent direction of allelic effect between SNPs associ-
ated with AD risk and plasma lipid levels (LDL, HDL, and tri-
glycerides). Additionally, for HS3ST1 and ECHD3 variants, we
found an opposite direction of allelic effect between increased
AD risk and triglycerides levels. One hypothesis for these find-
ings is that the observed pleiotropy between AD and plasma
lipids could be attributable to different haplotypes/gene alleles
involving the same SNPs. Another equally plausible hypothesis
is that the same haplotypes/gene alleles are involved in both AD
and plasma lipids but the underlying biological mechanisms are
distinct. From these findings, it seems less likely that the pleio-
tropic SNPs detected in this study influence AD pathogenesis
via cholesterol-mediated pathways.

Unlike epidemiological studies, coheritability analyses or bivariate GWAS methods, one strength of our present
approach is the ability to detect genetic pleiotropy even when there is no correlation of the signed effects (mixed direction-ality of effect). The conditional FDR method can detect SNPs that have a nonnull effect in one trait and that also tend to have a nonnull effect in another trait, independently of direction-ality. Another strength of this framework is leveraging genetic signal in one phenotype to identify variants in a second phe-
notype that would otherwise not be detected with a single phe-
notype approach. We note that the conditional FDR approach
allows reordering (and reranking) of SNPs based on P value
significance in the second phenotype (eg, CRP or triglycer-
ides), thus enabling identification of novel SNPs in the primary phenotype (eg, AD). In addition, as previously demonstrated,
these genetic analysis methods result in improved sensitivity
for a given specificity. Using this “pleiotropic” approach, we
detected 55 novel variants, indicating that genetic enrichment
improves statistical power for gene discovery.

In meta-analyses, we discovered 2 GWAS significant
AD susceptibility loci. The closest genes associated with

Figure 3. Forest plots for (A) rs13113697 on chromosome 4 and (B) rs7920721 on chromosome 10. CI indicates confidence interval; IGAP, International Genomics of Alzheimer’s Disease Project; and OR, odds ratio.

Figure 4. Regional association plots for (A) rs13113697 on chromosome 4 and (B) rs7920721 on chromosome 10. Linkage disequilibrium measured in the 1000 genomes European populations with plink version 1.07.
the 2 risk variants showed altered RNA levels in postmortem AD brains compared with control brains, suggesting a functional role. The first variant (rs13113697) is closest to the HS3ST1 gene on chromosome 4 (Figure 4A), which encodes heparan sulfate glucosaminyl 3-O-sulfotransferase, an intraluminal Golgi protein enzyme with multiple biological activities.31 The second variant (rs7920721) is closest to the ECHDC3 gene on chromosome 10 (Figure 4B), which encodes an enzyme called enoyl CoA hydratase domain containing 3.32 We note that by conditioning on cardiovascular traits and evaluating additional AD cohorts (deCODE and DemGene), we were able to find genome-wide significant evidence for previously33 suggested signal close to HS3ST1 and ECHDC3. At a value of \( P < 1.0 \times 10^{-6} \), we additionally found 3 suggestive variants on chromosome 11 (rs7396366; closest gene, APA2A), chromosome 15 (rs3131609; closest gene, USP50), and chromosome 17 (rs2526378; closest gene, BZRAP1).

It is important to note that in this study the diagnosis of AD was established clinically. Postmortem evidence from community- and population-based cohorts indicates that vascular brain injury often presents concomitantly with AD pathology and correlates with cognitive impairment above and beyond AD neuropathology.34 It is feasible that the individuals with clinically diagnosed AD from the IGAP, deCODE, and DemGene cohorts may have concomitant vascular brain disease, which may further contribute to their cognitive decline and dementia. Therefore, an alternative interpretation of our findings is that the susceptibility loci identified in this study may increase brain vulnerability to vascular or inflammatory insults, which in turn may exacerbate the clinical consequences of AD pathological changes.

### Conclusions

We found polygenic overlap among AD, CRP, and plasma lipids and leveraged this association to identify 2 novel genome-wide significant variants associated with increased AD risk. Careful and considerable effort will be required to further characterize the novel candidate genes detected in this study and to detect the functional variants responsible for the association of these loci with AD risk. Although no single common variant may be informative clinically, a combination of variants involved with inflammation or lipid metabolism may help identify older individuals at increased risk for AD. Our findings may also have implications for AD prevention trials involving anti-inflammatory agents.

### Acknowledgments

We thank the IGAP for providing summary results data for these analyses.

### Sources of Funding

This research was supported by grants from the National Institutes of Health (K02 NS067427, T32 EB005970, R01GM104400-01A, R01MH100351, AG033193, and U0149505), the Research Council of Norway (grants 213837, 225989, 232723, 237250/EU JPND), the South East Norway Health Authority (2013–123), the Norwegian Health Association, and the KG Jebsen Foundation. A.J. Schork was supported by National Institutes of Health grants RC2DA029475 and R01HD061414 and the Robert J. Glushko and Pamela Samuelson Graduate Fellowship. Dr Dehghan was supported by an NWO grant (veni, 916.12.154) and the Erasmus University Rotterdam Fellowship. Please see the Acknowledgments in the only online Data Supplement for IGAP funding sources.

### Disclosures

None.

### References


CLINICAL PERSPECTIVE

Late-onset Alzheimer disease (AD), the most common form of dementia, places a large economic and financial burden on families and society. With the aging of the US population and high costs associated with caring for cognitively impaired elderly, identifying strategies that prevent AD is of utmost importance. Cardiovascular disease (CVD) is being increasingly recognized as an important etiologic characteristic of AD. Many CVD traits such as dyslipidemia and inflammation can serve as therapeutic targets and are modifiable. Thus, focusing on the genetic and molecular overlap between CVD and AD may provide avenues to prevent or delay AD pathology. Although a number of epidemiological studies have examined the association between AD and CVD, a significant limitation of this approach is the inability to determine causal associations between CVD risk factors and AD pathobiology. Large-scale genome-wide association studies provide valuable insights into the role of specific biological pathways in disease pathogenesis, and combining genome-wide association studies from multiple disorders and phenotypes could elucidate shared pathobiology. Using data from large recent genome-wide association studies, we examined genetic overlap between AD, dyslipidemia, and inflammation. We found that genetic variants associated with increased plasma lipid and C-reactive protein levels are also associated with increased AD risk. By conditioning on genetic variants associated with the cardiovascular traits, we found novel AD susceptibility loci, including 2 genome-wide significant variants conferring increased risk for AD. Our findings provide novel insights into the relationship among AD pathogenesis, inflammation, and dyslipidemia and may have implications for AD prevention trials involving anti-inflammatory agents.