

# Intermediate Phenotypes Identify Divergent Pathways to Alzheimer's Disease

Joshua M. Shulman<sup>1,2,3</sup>, Lori B. Chibnik<sup>1,2,3</sup>, Cristin Aubin<sup>1,2,3</sup>, Julie A. Schneider<sup>4,5</sup>, David A. Bennett<sup>5</sup>, Philip L. De Jager<sup>1,2,3\*</sup>

**1** Program in Translational NeuroPsychiatric Genomics, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **2** Harvard Medical School, Boston, Massachusetts, United States of America, **3** Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America, **4** Department of Pathology, Rush University Medical Center, Chicago, Illinois, United States of America, **5** Rush Alzheimer's Disease Center, Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, United States of America

## Abstract

**Background:** Recent genetic studies have identified a growing number of loci with suggestive evidence of association with susceptibility to Alzheimer's disease (AD). However, little is known of the role of these candidate genes in influencing intermediate phenotypes associated with a diagnosis of AD, including cognitive decline or AD neuropathologic burden.

**Methods/Principal Findings:** Thirty-two single nucleotide polymorphisms (SNPs) previously implicated in AD susceptibility were genotyped in 414 subjects with both annual clinical evaluation and completed brain autopsies from the Religious Orders Study and the Rush Memory and Aging Project. Regression analyses evaluated the relation of SNP genotypes to continuous measures of AD neuropathology and cognitive function proximate to death. A SNP in the zinc finger protein 224 gene (*ZNF224*, rs3746319) was associated with both global AD neuropathology ( $p = 0.009$ ) and global cognition ( $p = 0.002$ ); whereas, a SNP at the phosphoenolpyruvate carboxykinase locus (*PCK1*, rs8192708) was selectively associated with global cognition ( $p = 3.57 \times 10^{-4}$ ). The association of *ZNF224* with cognitive impairment was mediated by neurofibrillary tangles, whereas *PCK1* largely influenced cognition independent of AD pathology, as well as Lewy bodies and infarcts.

**Conclusions/Significance:** The findings support the association of several loci with AD, and suggest how intermediate phenotypes can enhance analysis of susceptibility loci in this complex genetic disorder.

**Citation:** Shulman JM, Chibnik LB, Aubin C, Schneider JA, Bennett DA, et al. (2010) Intermediate Phenotypes Identify Divergent Pathways to Alzheimer's Disease. PLoS ONE 5(6): e11244. doi:10.1371/journal.pone.0011244

**Editor:** Katharina Domschke, University of Muenster, Germany

**Received:** April 25, 2010; **Accepted:** June 1, 2010; **Published:** June 21, 2010

**Copyright:** © 2010 Shulman et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** JMS is supported by NIH grant K08AG034290 and by the Clinical Investigator Training Program: Beth Israel Deaconess Medical Center and Harvard/MIT Health Sciences and Technology, in collaboration with Pfizer Inc. and Merck & Co. The Religious Orders Study and the Rush Memory and Aging Project were supported by NIH grants P30AG10161, R01AG15819, R01AG17917. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** JMS is supported by the Clinical Investigator Training Program: Beth Israel Deaconess Medical Center and Harvard/MIT Health Sciences and Technology, in collaboration with Pfizer Inc. and Merck & Co. This fellowship program is administered by academic institutions (Harvard/MIT/BIDMC), and supported by unrestricted commercial grants. This does not alter the authors' adherence to PLoS ONE policies on sharing data and materials. There are no other competing interests to declare.

\* E-mail: pdejager@rics.bwh.harvard.edu

## Introduction

Alzheimer's disease (AD), the most common cause of dementia, leads to progressive loss of memory and other cognitive domains, and is characterized pathologically by the accumulation of extracellular amyloid plaques and intracellular neurofibrillary tangles. AD likely develops from an interaction of numerous genes along with environmental risk factors, each with modest and incompletely penetrant effects. Linkage studies have identified rare gene mutations as causal in familial, early age-of-onset AD, but these Mendelian variants only explain a small fraction of disease burden in the general population [1]. The identification of susceptibility loci for sporadic, late age-of-onset AD has been more challenging, with numerous reports of candidate gene associations, most of which have not been consistently replicated in follow-up studies [2–4]. One notable exception is the *apolipoprotein E* locus (*APOE*): the  $\epsilon 4$  allele is common, increases AD susceptibility 3-fold, and is estimated to explain at least 10% of the population-

attributable risk of disease [1]. In addition, the *APOE*  $\epsilon 2$  allele is a validated AD protective allele though it is less common, and its effect size is more modest than  $\epsilon 4$ .

Genome-wide association (GWA) studies have emerged as a promising approach to identify susceptibility loci in common diseases with complex genetic inheritance, but until recently, most GWA scans in AD have been relatively underpowered, and identified loci have not been consistently replicated [5–14]. Increasing sample size is one approach for boosting statistical power, and this strategy has recently led to the identification of several promising new AD susceptibility loci, including *CRI*, *CLU*, and *PICALM* [15,16]. However, clinical heterogeneity remains a significant confounder of the case/control study design in AD, due to the likely inclusion of dementia cases with multiple pathologies, such as cerebrovascular disease or other neurodegenerative conditions. In addition, since AD develops following a protracted pre-clinical phase consisting of mild symptoms, control groups are susceptible to contamination by latent disease cases. Substantial

AD pathology is often present in advanced age, including in those with minimal or no cognitive impairment at death [17]. Subjects with significant pathology but subclinical disease are likely to dilute power in an AD case/control association analysis.

One approach to overcoming these obstacles is to study quantitative intermediate phenotypes. The manifestation of the AD clinical syndrome is the final culmination of a sequence of events beginning with genetic and environmental risk factors that trigger intermediate pathological changes, synapse loss and cell death, and ultimately cognitive decline and dementia. Outcome measures selected more proximally along this causal chain are expected to be less confounded and more strongly associated with susceptibility loci. In addition, compared to the dichotomous clinical diagnosis, quantitative intermediate phenotypes can capture more of the underlying heritable trait variation, further enhancing statistical power. Based on this promise, a number of studies have begun to take advantage of intermediate phenotypes for genetic association analysis in AD, including neuropsychiatric test measures [18], MRI imaging data [19,20], biomarkers from blood and CSF [21,22], and direct measurements of AD pathology [23]. The latter approach requires access to large study populations with detailed clinical and neuropathologic characterization. The Religious Orders Study and Rush Memory and Aging Project are prospectively following more than 2,300 older persons, all of whom have agreed to annual clinical evaluation and brain donation at death. More than 800 autopsies have been completed to date, and quantitative analyses of amyloid and tangle burden has been performed on nearly 600. In a recent study of *APOE* in this cohort, we found that intermediate cognitive and pathological phenotypes substantially increase power for genetic association analysis [23]. In addition, using neuropathologic phenotypes, the association between *APOE* and cognitive impairment was previously shown to be mediated by a sequential cascade of amyloid plaque formation and subsequent development of neurofibrillary tangle pathology [24,25]. Therefore, beyond enhancing power for association analysis, intermediate phenotypes hold the additional promise of testing mechanistic hypotheses of gene action.

In this study, we extend our previous work to evaluate several candidate AD susceptibility loci for associations with intermediate phenotypes relevant to AD. Thirty-two candidate SNPs were selected based on their discovery in AD GWA studies and/or evidence from the AlzGene online meta-analyses [2,26]. SNPs were genotyped in more than 400 subjects with detailed cognitive and pathological data, allowing assessment of genotype relations to quantitative AD pathology and cognitive function proximate to death. We subsequently leveraged the detailed phenotypes available in our cohorts to dissect the functional pathways that link genetic variants to cognitive impairment.

## Methods

### Ethics Statement

Written informed consent and an Anatomic Gift Act were signed by all Religious Orders Study and Rush Memory and Aging project participants after the procedures were fully explained, and both studies were approved by the Institutional Review Board of Rush University Medical Center. The work described in this report was additionally approved by the Institutional Review Boards of the Brigham and Women's Hospital and Massachusetts Institute of Technology.

### Subjects

Clinical and post-mortem data came from participants in the Religious Orders Study and Rush Memory and Aging Project, two

longitudinal, epidemiologic clinical-pathologic studies of aging and AD [17]. In both studies, participants without known dementia at baseline agreed to annual detailed clinical evaluation and brain donation at the time of death. Participants in the Religious Orders Study were older Catholic nuns, priests and brothers from about 40 groups in 12 states across the United States. Subjects in the Rush Memory and Aging Project were older, community-dwelling persons from about 40 retirement communities and subsidized senior housing facilities across northeastern Illinois. Since 1993, more than 2,300 persons agreed to participate in these studies. The overall follow-up rate exceeds 90% of survivors and the overall autopsy rate exceeds 90% of decedents. Of those subjects with completed neuropathologic analyses, and following genotyping quality control filters, 414 persons with genotyping data were available for analysis in February of 2009 when this study was initiated (250 from the Religious Orders Study and 164 from the Rush Memory and Aging Project).

### Clinical evaluation

The clinical diagnoses of dementia and AD were made each year following the recommendations of the joint working group of the National Institute of Neurologic and Communicative Disorders and Stroke and the AD and Related Disorders Association [27], as previously described in detail [28]. Probable AD refers to persons with clinical AD and no other clinical condition contributing to cognitive impairment and possible AD refers to persons meeting inclusion criteria for AD who are thought to have another condition (e.g., stroke) contributing to cognitive impairment. MCI referred to those individuals rated as cognitively impaired by the neuropsychologist but not demented by the examining physician, as previously described [29]. At the time of death, clinical data were reviewed by a neurologist without access to post-mortem data and a summary diagnostic opinion was rendered regarding the most likely clinical diagnosis at the time of death. Level of cognition was based on cognitive testing performed proximate to death. The Religious Orders Study and Rush Memory and Aging Project have 19 cognitive performance tests in common, and use identical analytic procedures to develop summary statistics. Mini-Mental State Examination [30] was used to describe the cohort and one test was used for diagnostic classification purposes only. The remaining 17 tests have been previously described [17]. Tests were converted to *z* scores, using the mean and SD from the baseline evaluation of all participants, and averaged to yield summary measures of global cognition and five cognitive domains: episodic memory, semantic memory, working memory, perceptual speed, and visuospatial ability. Summary measures minimize floor and ceiling effects and other sources of random variability. For the mediation analyses incorporating diagnosis of diabetes, annual clinical evaluations allowed documentation of history of diabetes and use of medications to treat diabetes. Diabetes was determined to be present if the participant was ever taking a medication, such as insulin or an oral hypoglycemic, to treat diabetes, as determined by direct inspection of medication containers, or ever reported a history of diagnosis of diabetes, or both, as previously described [31].

### Neuropathological evaluation

Brain autopsies were performed across the US as previously described [17]. Bielschowsky silver stain was used to visualize neuritic plaques, diffuse plaques, and neurofibrillary tangles in tissue sections from the midfrontal, middle temporal, inferior parietal, and entorhinal cortices, and the hippocampal CA1 sector. The neuropathologic diagnosis of AD was made by a board-

certified neuropathologist without access to any clinical data as previously reported [17,28]. We classified persons as having pathologic AD based on intermediate or high likelihood of AD by National Institute on Aging (NIA)-Reagan criteria using CERAD estimates of neuritic plaque density and Braak staging of neurofibrillary pathology [32–34], as previously described [17]. The quantitative composite AD pathology score was based on counts of neuritic plaques, diffuse plaques and neurofibrillary tangles as previously described [35,36]. Because the means, standard deviations, and ranges of the data varied widely for the pathologic indices, we converted the raw counts to a standard distribution by dividing each person's count by the standard deviation for that particular count and formed a summary measure by averaging the scaled scores. Because the data were skewed, square root of the scaled score was used in analyses. Separate summary measures of neurofibrillary tangles and neuritic and diffuse plaques were also made. Chronic macroscopic cerebral infarctions and alpha-synuclein immunoreactive Lewy bodies were determined as previously described and considered present or absent for analyses [17].

### Genotyping

DNA was extracted from lymphocytes or frozen post-mortem brain tissue. *APOE* genotyping was performed by Agencourt Bioscience Corporation (Beverly, MA) utilizing high throughput sequencing of codon 112 (position 3937) and codon 158 (position 4075) of exon 4 of the *APOE* gene on chromosome 19. In addition to the *APOE*  $\epsilon 4$  and  $\epsilon 2$  alleles, 32 SNPs were selected for genotyping in our cohort, based on prior evidence from the literature, as of February, 2009. Thus, the more recently discovered *CRI*, *CLU*, and *PICALM* [15,16] loci were not included in this study, but are the focus of a separate study (Chibnik et al., submitted). The selected SNPs were equally divided between the top results of AD case/control GWA studies [5–7,9,10–14] (16 SNPs) and candidate gene association studies (16 SNPs), which were chosen based on their top ranking in AlzGene meta-analyses [2]. The 32 candidate SNPs were genotyped using matrix-assisted laser desorption-ionization time-of-flight mass spectrometry on a MassARRAY platform (Sequenom). After excluding subjects for failed genotyping exceeding the 10% threshold, 414 individuals remained for subsequent analysis (genotyping rate in these subjects was >99%). All SNP allele frequencies satisfied Hardy-Weinberg equilibrium ( $p > 0.001$ ). Allele frequencies were not significantly different between the Religious Orders Study and Rush Memory and Aging Project subjects, supporting the validity of pooled analyses.

### Statistical Analysis

Given the complementary study designs and similar procedures for data collection and generation of the cognitive and neuropathologic outcome traits, we pooled data from the Religious Orders Study and Rush Memory and Aging Project for our analyses, consistent with numerous prior studies [17,23]. Genetic association was performed using the PLINK analysis software toolkit [37]. Linear regression was used to evaluate the association of allele genotypes with level of cognition proximate to death in a 2-degree-of-freedom, genotypic test of association, with covariates included for age, gender, and years of education. In order to refine the genetic model, we additionally tested selected SNPs using a 1-degree-of-freedom test to examine for additive, dominant, or recessive allelic effects. These studies were performed using PLINK as well as the R statistical computing program (www.r-project.org). Linear regression modeling in R was used to calculate residual quantitative trait variance explained, and to perform

statistical mediation analyses. For the case/control association analysis based on AD clinical diagnosis, logistic regression was performed in PLINK under both additive and genotypic models, and again including covariates for age, gender, and education. All  $p$ -values reported are unadjusted for multiple hypothesis testing. A Bonferroni-corrected significance threshold of  $p < 0.001$  was calculated for the 34 SNPs tested for associations with our two primary outcomes, a quantitative measure of global cognitive performance and global AD pathology. Given the high correlation between the pathologic and cognitive traits, applying an adjustment for 68 tests would be overly conservative. Otherwise, the threshold of  $p < 0.01$  was selected to indicate suggestive statistical evidence of association. All other evaluated phenotypes, including those for AD clinical diagnosis, pathology sub-types, and cognitive domains were considered secondary analyses.

## Results

### Associations with global AD pathology and global cognition

Subject demographics, clinical and neuropathologic diagnoses, cognitive status, and *APOE* genotypes for the cohort analyzed in this study are presented in Table 1. In clinical evaluations proximate to death, of the 414 subjects in our study cohort, 131 (31.6%) had normal cognition, 98 (23.7%) had mild cognitive impairment, and 185 (44.7%) were demented (173 met criteria for possible or probable AD). As expected, a significant proportion (41.5%) of individuals without dementia satisfied NIA-Reagan pathological criteria for intermediate or high likelihood AD.

We initially tested for associations between each of the 34 polymorphisms and our two primary outcomes, intermediate phenotypes representing a measure of global AD pathologic burden on autopsy and a measure of global cognitive function proximate to death (Table 2). Linear regression models were used to examine the relation of SNP genotypes to the quantitative neuropathologic and cognitive traits in a 2 degree-of-freedom statistical test, adjusting for the effects of age at death, gender, and years of education. As expected, *APOE*  $\epsilon 4$  was significantly associated with both cognition ( $p = 3.4 \times 10^{-10}$ ) and AD pathology ( $p = 1.6 \times 10^{-24}$ ) in our cohort, whereas an association with *APOE*  $\epsilon 2$  was only seen for the pathological phenotype ( $p = 9.1 \times 10^{-4}$ ). In addition, we found associations with AD intermediate phenotypes

**Table 1.** Demographic, clinical and pathological characteristics of the study cohort.

<b>n</b>	414
<b>Mean age, y (SD)</b>	87.1 (6.9)
<b>Male (%)</b>	161 (38.9)
<b>Education, y (SD)</b>	16.5 (3.6)
<b>Mini-Mental State Examination (SD)</b>	21.4 (9.2)
<b>NINCDs clinical AD (%)</b>	173 (41.8)
<b>Mild cognitive impairment (%)</b>	98 (23.7)
<b>Diabetes (%)</b>	78 (18.8)
<b>APOE <math>\epsilon 4</math> allele present (%)</b>	121 (29.2)
<b>APOE <math>\epsilon 2</math> allele present (%)</b>	64 (15.5)
<b>NIA-Reagan pathological AD (%)</b>	236 (57.6)
<b>Lewy bodies (%)</b>	77 (18.6)
<b>Infarcts (%)</b>	143 (34.5)

doi:10.1371/journal.pone.0011244.t001

**Table 2.** Relation of candidate AD polymorphisms to intermediate cognitive and pathologic phenotypes.

Chr	SNP	Gene	Reference <sup>1</sup>	Alleles		Global Pathology (p <sup>3</sup> )	Global Cognition (p <sup>3</sup> )
				Minor/Major	MAF <sup>2</sup>		
1	<i>Rs4845378</i>	nicotinic cholinergic receptor ( <i>CHRNA2</i> )	AlzGene	T/G	0.09	0.794	0.475
	<i>Rs505058</i>	lamin A/C ( <i>LMNA</i> )	Grube et al. 2007	C/T	0.10	0.366	0.606
	<i>Rs2584820</i>	regulator of G protein signaling protein-like 2 ( <i>RGS2</i> )	Liu et al. 2007	G/A	0.05	0.823	0.298
	<i>Rs12044355</i>	disruptend in schizophrenia 1 ( <i>DISC1</i> )	Beecham et al. 2009	C/A	0.35	0.959	0.296
2	<i>Rs1800587</i>	interleukin-1 Alpha ( <i>IL1A</i> )	AlzGene	A/G	0.31	0.371	0.724
	<i>Rs1143634</i>	interleukin-1 Beta ( <i>IL1B</i> )	AlzGene	A/G	0.21	0.392	0.907
3	<i>Rs1049296</i>	transferrin ( <i>TF</i> )	AlzGene	T/C	0.15	0.245	0.945
4	<i>Rs727153</i>	lecithin retinol acyltransferase ( <i>LRAT</i> )	Abraham et al. 2008	C/T	0.48	0.157	0.911
9	<i>Rs7019241</i>	golgi membrane protein 1 ( <i>GOLM1</i> )	Li et al. 2008	T/C	0.12	0.512	0.705
	<i>Rs10868366</i>	golgi membrane protein 1 ( <i>GOLM1</i> )	Li et al. 2008	T/G	0.12	0.490	0.730
	<i>Rs4878104</i>	death associated protein kinase ( <i>DAPK1</i> )	Li et al. 2006	T/C	0.37	0.508	0.384
10	<i>Rs2306604</i>	transcription factor A, mitochondrial ( <i>TFAM</i> )	AlzGene	G/A	0.47	0.503	0.742
	<i>Rs13500</i>	cholesterol 25-hydroxylase ( <i>CH25H</i> )	AlzGene	T/C	0.10	0.368	0.579
	<i>Rs2986017</i>	calcium homeostasis modulator 1 ( <i>CALHM1</i> )	AlzGene	A/G	0.26	0.231	0.505
	<i>Rs600879</i>	sortilin-related VPS10-containing receptor ( <i>SORCS1</i> )	AlzGene	A/G	0.11	0.290	0.849
	<i>Rs1903908</i>	CG2039140	Grube et al. 2007	A/G	0.15	0.846	0.912
11	<i>Rs6265</i>	brain derived neurotrophic factor ( <i>BDNF</i> )	AlzGene	A/G	0.19	0.032	0.347
	<i>Rs1385600</i>	GRB2-associated binding protein 2 ( <i>GAB2</i> )	Reiman et al. 2007	C/T	0.18	0.825	0.966
	<i>Rs2373115</i>	GRB2-associated binding protein 2 ( <i>GAB2</i> )	Reiman et al. 2007	T/G	0.18	0.544	0.981
	<i>Rs2070045</i>	sortilin-related receptor ( <i>SORL1</i> )	AlzGene	G/T	0.22	0.182	0.453
	<i>Rs3824968</i>	sortilin-related receptor ( <i>SORL1</i> )	AlzGene	A/T	0.30	0.831	0.683
12	<i>Rs11610206</i>	FAM113B	Beecham et al. 2009	C/T	0.08	0.227	0.519
14	<i>Rs11159647</i>	14q31.2	Bertram et al. 2008	A/G	0.49	0.968	0.581
17	<i>Rs1554948</i>	tyrosine kinase non-receptor 1 ( <i>TNKS1</i> )	Grube et al. 2007	A/T	0.46	0.441	0.337
	<i>Rs2471738</i>	microtubule associated protein tau ( <i>MAPT</i> )	AlzGene	T/C	0.21	0.533	0.392
	<i>Rs1800764</i>	angiotensin converting enzyme ( <i>ACE</i> )	AlzGene	C/T	0.47	0.451	0.489
19	<i>Rs3746319</i>	zinc finger protein 224 ( <i>ZNF224</i> )	Beecham et al. 2009	A/G	0.16	<b>0.009</b>	<b>0.004</b>
	<i>Rs3826656</i>	CD33	Bertram et al. 2008	G/A	0.20	0.258	0.455
	<i>Rs3745833</i>	galinin-related receptor ( <i>GALP</i> )	Grube et al. 2007	C/A	0.36	0.024	0.288
20	<i>Rs1799990</i>	prion protein ( <i>PRNP</i> )	AlzGene	G/A	0.33	0.727	0.800
	<i>Rs8192708</i>	phosphoenolpyruvate carboxykinase 1 ( <i>PCK1</i> )	Grube et al. 2007	G/A	0.12	0.056	<b>3.57 × 10<sup>-4</sup></b>
23	<i>Rs5984894</i>	protocadherin 11 X-linked ( <i>PCDH11X</i> )	Carrasquillo et al. 2009	A/G	0.48	0.188	0.624
19	$\epsilon 4$	apolipoprotein E ( <i>APOE</i> )	-	-	0.16	<b>1.56 × 10<sup>-24</sup></b>	<b>3.40 × 10<sup>-10</sup></b>
	$\epsilon 2$	apolipoprotein E ( <i>APOE</i> )	-	-	0.08	<b>9.15 × 10<sup>-4</sup></b>	0.238

1 SNPs were selected based on AlzGene meta-analyses (ref. 2) or from results of AD GWA studies (refs. 5–7, 9–14).

2 MAF = minor allele frequency.

3 Unadjusted p-values from genotypic regression models, including covariates for age, gender, and years of education.

doi:10.1371/journal.pone.0011244.t002

for two SNPs, within the zinc finger protein 224 (*ZNF224*) and phosphoenolpyruvate carboxykinase 1 (*PCK1*) genes, both of which were selected for genotyping based on their identification in AD case/control GWA studies [6,10]. The *ZNF224* SNP (*rs3746319*) was associated with both global cognition ( $p = 0.009$ ) and global AD pathology ( $p = 0.004$ ). In contrast, the *PCK1* SNP (*rs8192708*) was significantly associated with global cognition ( $p = 3.57 \times 10^{-4}$ ) but not global AD pathology ( $p = 0.056$ ), suggesting that this locus may influence cognitive impairment through mechanisms other than AD pathology. Besides *APOE*  $\epsilon 4$ , none of

the SNP associations surpass the currently accepted threshold for genome-wide significance ( $p < 5.0 \times 10^{-8}$ ); however, the association between *PCK1* and global cognition exceeds a Bonferroni-corrected significance threshold of  $p < 0.001$  for 34 independent tests. Given the high correlation between the pathologic and cognitive traits, applying an adjustment for 68 tests would be overly conservative; however, the *PCK1* association still exceeds that standard ( $p < 7 \times 10^{-4}$ ).

Although the risk alleles for the associations of both the *ZNF224* and *PCK1* loci with the intermediate phenotypes in our cohort also

increase risk of AD diagnosis (Table 3), their effects are opposite to that reported in the original GWA studies [6,10]. In the case of *ZNF224*, we find that the minor allele, *rs3746319*<sup>4</sup>, is associated with both increased AD pathologic burden and decreased cognitive performance; whereas this variant was protective against AD in the GWA study (G. Beecham and M. Pericak-Vance, personal communication). Similarly, for *PCK1*, the minor allele, *rs8192708*<sup>4</sup>, significantly protected against cognitive decline in our cohort but was in fact the AD risk allele in the original GWA study [2]. Interestingly, two subsequent replication analyses of *rs8192708* documented associations of decreased AD risk with the minor allele, consistent with our findings [4,38]. Therefore, while the

effects of the *ZNF224* and *PCK1* loci on a diagnosis of AD and on intermediate phenotypes are consistent within our study (and in two other *PCK1* replication studies); they are not consistent with the original GWA analyses. In the discussion section, we further address possible explanations for these discrepancies.

### Associations with neuritic and diffuse plaques, neurofibrillary tangles, and cognitive subdomains

The global AD pathology score averages the post-mortem density of neuritic and diffuse plaques and neurofibrillary tangles in multiple brain regions; however, we hypothesized that certain

**Table 3.** Relation of candidate AD polymorphisms to clinical AD diagnosis.

SNP	Gene	A <sup>1</sup>	MAF <sup>2</sup>		OR (95% CI) <sup>3</sup>	p-value <sup>4</sup>	
			AD cases	controls		additive	genotypic
<i>rs4845378</i>	<i>CHRN2</i>	T	0.08	0.10	0.77 (0.46–1.28)	0.305	0.509
<i>rs505058</i>	<i>LMNA</i>	C	0.08	0.11	0.81 (0.50–1.33)	0.411	0.571
<i>rs2584820</i>	<i>RGSL2</i>	G	0.04	0.06	0.63 (0.31–1.26)	0.191	0.589
<i>rs12044355</i>	<i>DISC1</i>	C	0.35	0.35	1.07 (0.80–1.44)	0.641	0.716
<i>rs1800587</i>	<i>IL1A</i>	A	0.30	0.31	0.98 (0.72–1.34)	0.896	0.901
<i>rs1143634</i>	<i>IL1B</i>	A	0.19	0.22	0.87 (0.61–1.26)	0.466	0.729
<i>rs1049296</i>	<i>TF</i>	T	0.13	0.17	0.76 (0.51–1.14)	0.183	0.285
<i>rs727153</i>	<i>LRAT</i>	C	0.48	0.49	1.03 (0.77–1.37)	0.860	0.803
<i>rs7019241</i>	<i>GOLM1</i>	T	0.12	0.12	0.96 (0.60–1.54)	0.871	0.786
<i>rs10868366</i>	<i>GOLM1</i>	T	0.13	0.13	0.94 (0.59–1.50)	0.804	0.761
<i>rs4878104</i>	<i>DAPK1</i>	T	0.37	0.37	0.98 (0.74–1.31)	0.907	0.984
<i>rs2306604</i>	<i>TFAM</i>	G	0.48	0.47	1.01 (0.75–1.36)	0.954	0.354
<i>rs13500</i>	<i>CH25H</i>	T	0.10	0.10	0.92 (0.58–1.47)	0.731	0.784
<i>rs2986017</i>	<i>CALHM1</i>	A	0.27	0.26	1.08 (0.78–1.48)	0.645	0.827
<i>rs600879</i>	<i>SORCS1</i>	A	0.11	0.11	0.91 (0.56–1.45)	0.683	0.71
<i>rs1903908</i>	<i>CG2039140</i>	A	0.16	0.14	1.14 (0.76–1.71)	0.525	0.707
<i>rs6265</i>	<i>BDNF</i>	A	0.19	0.18	1.03 (0.71–1.50)	0.885	0.83
<i>rs1385600</i>	<i>GAB2</i>	C	0.18	0.18	1.03 (0.70–1.51)	0.897	0.73
<i>rs2373115</i>	<i>GAB2</i>	T	0.18	0.18	0.98 (0.66–1.45)	0.913	0.909
<i>rs2070045</i>	<i>SORL1</i>	G	0.21	0.22	0.96 (0.67–1.39)	0.846	0.522
<i>rs3824968</i>	<i>SORL1</i>	A	0.31	0.30	0.98 (0.71–1.35)	0.881	0.891
<i>rs11610206</i>	<i>FAM113B</i>	C	0.08	0.08	1.02 (0.58–1.79)	0.946	0.985
<i>rs11159647</i>	<i>14q31.2</i>	A	0.51	0.47	1.17 (0.87–1.57)	0.304	0.588
<i>rs1554948</i>	<i>TNK1</i>	A	0.46	0.46	0.91 (0.68–1.22)	0.519	0.727
<i>rs2471738</i>	<i>MAPT</i>	T	0.22	0.20	1.10 (0.77–1.57)	0.606	0.773
<i>rs1800764</i>	<i>ACE</i>	C	0.49	0.47	1.06 (0.80–1.42)	0.673	0.507
<i>rs3746319</i>	<i>ZNF224</i>	A	0.19	0.15	1.51 (1.02–2.25)	0.042	<b>0.008</b>
<i>rs3826656</i>	<i>CD33</i>	G	0.20	0.21	0.93 (0.64–1.33)	0.682	0.84
<i>rs3745833</i>	<i>GALP</i>	C	0.36	0.36	0.94 (0.69–1.28)	0.699	0.845
<i>rs1799990</i>	<i>PRNP</i>	G	0.32	0.33	0.97 (0.71–1.33)	0.844	0.906
<i>rs8192708</i>	<i>PCK1</i>	G	0.09	0.14	0.51 (0.32–0.82)	<b>0.005</b>	0.011
<i>rs5984894</i>	<i>PCDH11X</i>	A	0.43	0.52	0.74 (0.58–0.96)	0.021	0.068
<i>ε4</i>	<i>APOE</i>	-	0.21	0.11	2.67 (1.74–4.11)	<b>7.0 × 10<sup>-6</sup></b>	<b>9.63 × 10<sup>-6</sup></b>
<i>ε2</i>	<i>APOE</i>	-	0.08	0.08	0.76 (0.44–1.32)	0.336	0.876

1 A = minor allele.

2 MAF = minor allele frequency.

3 OR = odds ratio, CI = confidence interval.

4 Unadjusted p-values from logistic regression models, under both additive and genotypic models, including covariates for age, gender, and education.

doi:10.1371/journal.pone.0011244.t003

AD susceptibility loci might selectively promote one type of pathology, in which case the composite pathologic outcome might dilute statistical power to detect associations. We therefore performed secondary analyses to determine whether any of the candidate SNPs tested demonstrate selective or more robust association signals with separate quantitative measures of plaques or tangle pathology (Table 4). All analyses were again performed using linear regression models to test for associations with SNP genotypes, adjusted for the effects of age, gender, and education. A SNP at the *GALP* locus (*rs3745833*) showed suggestive evidence for association with diffuse plaques ( $p = 0.003$ ), but not with neurofibrillary tangles ( $p = 0.373$ ). In contrast, the  $\zeta$ *NF224* SNP (*rs3746319*) was strongly associated with neurofibrillary tangle

burden ( $p = 1.49 \times 10^{-4}$ ), whereas no significant association was seen with either neuritic plaque ( $p = 0.018$ ) or diffuse plaque ( $p = 0.290$ ) pathology. Therefore, the association with the tangle subscore is likely the primary driver for the  $\zeta$ *NF224* locus association with global AD pathology ( $p = 0.009$ ), and the composite score appears to dilute statistical power. Interestingly, the *PCK1* SNP (*rs8192708*), which was not associated with the global pathology measure, did show suggestive evidence for association with neuritic plaque pathology ( $p = 0.007$ ); however, this did not appear to explain the strong association with global cognition ( $p = 3.57 \times 10^{-4}$ ), as investigated further below.

Similar to the approach taken with the pathological phenotypes, we performed secondary analyses to assess whether any of the

**Table 4.** Relation of polymorphisms to amyloid plaques and neurofibrillary tangles.

SNP	Gene	Amyloid Pathology		Tau Pathology
		Diffuse Plaques	Neuritic Plaques	Neurofibrillary Tangles
		(p)	(p)	(p)
<i>rs4845378</i>	<i>CHRN2</i>	0.554	0.859	0.137
<i>rs505058</i>	<i>LMNA</i>	0.511	0.307	0.520
<i>rs2584820</i>	<i>RGSL2</i>	0.199	0.235	0.342
<i>rs12044355</i>	<i>DISC1</i>	0.801	0.700	0.715
<i>rs1800587</i>	<i>IL1A</i>	0.265	0.285	0.778
<i>rs1143634</i>	<i>IL1B</i>	0.367	0.222	0.720
<i>rs1049296</i>	<i>TF</i>	0.379	0.274	0.521
<i>rs727153</i>	<i>LRAT</i>	0.041	0.034	0.670
<i>rs7019241</i>	<i>GOLM1</i>	0.766	0.443	0.686
<i>rs10868366</i>	<i>GOLM1</i>	0.673	0.509	0.764
<i>rs4878104</i>	<i>DAPK1</i>	0.673	0.398	0.420
<i>rs2306604</i>	<i>TFAM</i>	0.345	0.231	0.088
<i>rs13500</i>	<i>CH25H</i>	0.288	0.349	0.787
<i>rs2986017</i>	<i>CALHM1</i>	0.748	0.135	0.109
<i>rs600879</i>	<i>SORCS1</i>	0.523	0.198	0.395
<i>rs1903908</i>	<i>CG2039140</i>	0.913	0.866	0.680
<i>rs6265</i>	<i>BDNF</i>	0.117	0.050	0.068
<i>rs1385600</i>	<i>GAB2</i>	0.571	0.580	0.388
<i>rs2373115</i>	<i>GAB2</i>	0.739	0.323	0.168
<i>rs2070045</i>	<i>SORL1</i>	0.052	0.304	0.918
<i>rs3824968</i>	<i>SORL1</i>	0.794	0.806	0.417
<i>rs11610206</i>	<i>FAM113B</i>	0.605	0.071	0.438
<i>rs11159647</i>	<i>14q31.2</i>	0.398	0.616	0.599
<i>rs1554948</i>	<i>TNK1</i>	0.655	0.254	0.474
<i>rs2471738</i>	<i>MAPT</i>	0.401	0.769	0.239
<i>rs1800764</i>	<i>ACE</i>	0.437	0.494	0.627
<i>rs3746319</i>	<i>ZNF224</i>	0.290	0.018	<b><math>1.49 \times 10^{-4}</math></b>
<i>rs3826656</i>	<i>CD33</i>	0.769	0.558	0.062
<i>rs3745833</i>	<i>GALP</i>	<b>0.003</b>	0.103	0.373
<i>rs1799990</i>	<i>PRNP</i>	0.618	0.674	0.801
<i>rs8192708</i>	<i>PCK1</i>	0.551	<b>0.007</b>	0.080
<i>rs5984894</i>	<i>PCDH11X</i>	0.654	0.166	0.070
$\epsilon 4$	<i>APOE</i>	<b><math>3.74 \times 10^{-16}</math></b>	<b><math>3.44 \times 10^{-20}</math></b>	<b><math>6.5 \times 10^{-15}</math></b>
$\epsilon 2$	<i>APOE</i>	<b><math>7.16 \times 10^{-4}</math></b>	<b>0.001</b>	0.097

doi:10.1371/journal.pone.0011244.t004



SNPs showed more robust association with the five cognitive subdomains that comprise the global cognition score. Linear regression was again used to test for association of each SNP with separate quantitative trait outcomes representing episodic memory, semantic memory, working memory, perceptual speed, and visuospatial ability (Table 5). Episodic memory impairment, the most characteristic cognitive deficit of AD, was associated with both the *ZNF224* locus ( $p=0.003$ ) and the *PCK1* locus ( $p=3.69\times 10^{-4}$ ). *ZNF224* was additionally associated with decline in visuospatial function ( $p=0.007$ ), and *PCK1* showed evidence for association with semantic memory impairment ( $p=0.001$ ).

### Divergent pathways from genes to cognitive impairment

For the SNPs at the *ZNF224* and *PCK1* loci, we performed additional linear regression analyses to refine the genetic model for

the relation with our intermediate phenotypes (additive, dominant, or recessive), better characterize the strength of the observed effects, and develop statistical models to test hypotheses about mechanistic pathways. Our core regression model, consisting of age at death, gender, and years of education, explained 3% and 7% of the variation in our pathological and cognitive traits, respectively. Using the optimal dominant model of inheritance, the *ZNF224* SNP (*rs3746319*) explained an additional 2% (Beta = 0.13,  $p=0.003$ ) of the residual variance in global AD pathology and 2.1% (Beta = -0.39,  $p=0.002$ ) of the variance in global cognition (Table 6). We next explored whether the effect of this locus on AD pathology might mediate its association with cognition (Table 7). When a term for global AD pathology was incorporated in our linear regression model, the magnitude of the association between *ZNF224* and global cognition was attenuated

**Table 5.** Relation of polymorphisms to measures of cognitive performance.

SNP	Gene	Episodic Memory	Semantic Memory	Working Memory	Perceptual Speed	Visuospatial Ability
		(p)	(p)	(p)	(p)	(p)
<i>rs4845378</i>	<i>CHRNA2</i>	0.315	0.384	0.634	0.959	0.818
<i>rs505058</i>	<i>LMNA</i>	0.413	0.651	0.624	0.972	0.744
<i>rs2584820</i>	<i>RGS2</i>	0.332	0.317	0.311	0.572	0.630
<i>rs12044355</i>	<i>DISC1</i>	0.548	0.592	0.223	0.700	0.299
<i>rs1800587</i>	<i>IL1A</i>	0.440	0.360	0.761	0.940	0.766
<i>rs1143634</i>	<i>IL1B</i>	0.913	0.588	0.978	0.945	0.520
<i>rs1049296</i>	<i>TF</i>	0.995	0.956	0.702	0.412	0.379
<i>rs727153</i>	<i>LRAT</i>	0.923	0.904	0.423	0.828	0.951
<i>rs7019241</i>	<i>GOLM1</i>	0.516	0.515	0.380	0.245	0.489
<i>rs10868366</i>	<i>GOLM1</i>	0.479	0.641	0.564	0.314	0.490
<i>rs4878104</i>	<i>DAPK1</i>	0.191	0.715	0.934	0.590	0.797
<i>rs2306604</i>	<i>TFAM</i>	0.877	0.554	0.444	0.560	0.126
<i>rs13500</i>	<i>CH25H</i>	0.868	0.407	0.170	0.733	0.143
<i>rs2986017</i>	<i>CALHM1</i>	0.231	0.881	0.594	0.784	0.816
<i>rs600879</i>	<i>SORCS1</i>	0.601	0.786	0.955	0.484	0.266
<i>rs1903908</i>	<i>CG2039140</i>	0.640	0.959	0.714	0.273	0.504
<i>rs6265</i>	<i>BDNF</i>	0.196	0.375	0.462	0.903	0.913
<i>rs1385600</i>	<i>GAB2</i>	0.760	0.319	0.865	0.624	0.968
<i>rs2373115</i>	<i>GAB2</i>	0.957	0.359	0.949	0.702	0.939
<i>rs2070045</i>	<i>SORL1</i>	0.443	0.952	0.746	0.170	0.357
<i>rs3824968</i>	<i>SORL1</i>	0.716	0.849	0.898	0.344	0.510
<i>rs11610206</i>	<i>FAM113B</i>	0.268	0.574	0.572	0.572	0.692
<i>rs11159647</i>	<i>14q31.2</i>	0.275	0.883	0.749	0.277	0.989
<i>rs1554948</i>	<i>TNK1</i>	0.494	0.469	0.459	0.278	0.162
<i>rs2471738</i>	<i>MAPT</i>	0.509	0.401	0.165	0.033	0.202
<i>rs1800764</i>	<i>ACE</i>	0.475	0.152	0.736	0.153	0.849
<i>rs3746319</i>	<i>ZNF224</i>	<b>0.003</b>	0.031	0.022	0.089	<b>0.007</b>
<i>rs3826656</i>	<i>CD33</i>	0.534	0.310	0.928	0.408	0.767
<i>rs3745833</i>	<i>GALP</i>	0.212	0.433	0.245	0.956	0.637
<i>rs1799990</i>	<i>PRNP</i>	0.652	0.952	0.877	0.982	0.650
<i>rs8192708</i>	<i>PCK1</i>	<b><math>3.69\times 10^{-4}</math></b>	<b>0.0010</b>	0.051	0.012	0.073
<i>rs5984894</i>	<i>PCDH11X</i>	0.769	0.799	0.460	0.722	0.877
$\epsilon 4$	<i>APOE</i>	<b><math>2.36\times 10^{-12}</math></b>	<b><math>1.22\times 10^{-7}</math></b>	<b><math>5.60\times 10^{-6}</math></b>	<b><math>1.22\times 10^{-4}</math></b>	<b><math>5.46\times 10^{-4}</math></b>
$\epsilon 2$	<i>APOE</i>	0.130	0.943	0.555	0.126	0.665

doi:10.1371/journal.pone.0011244.t005

**Table 6.** Detailed genotype-phenotype data and statistical modeling.

Variant	Minor Allele Homozygotes			Heterozygotes			Major Allele Homozygotes			Beta (SE)	p (Model) <sup>2</sup>	Variance Explained
	No. Subjects (Freq.)	Genotype	Mean Score (SD) <sup>1</sup>	No. Subjects (Freq.)	Genotype	Mean Score (SD) <sup>1</sup>	No. Subjects (Freq.)	Genotype	Mean Score (SD) <sup>1</sup>			
<b>Global Pathology</b>												
<i>rs3746319</i> (ZNF224)	11 (0.03)	AA	0.749 (0.486)	113 (0.28)	AG	0.811 (0.427)	290 (0.70)	GG	0.687 (0.398)	0.13 (0.04)	0.003 (D)	2
<i>APOE ε2</i>	2 (0.005)	-	0.388 (0.010)	62 (0.15)	-	0.576 (0.409)	340 (0.84)	-	0.754 (0.409)	-0.20 (0.05)	$2.11 \times 10^{-4}$ (A)	3.1
<i>APOE ε4</i>	9 (0.02)	-	0.906 (0.437)	110 (0.27)	-	1.011 (0.357)	285 (0.71)	-	0.609 (0.376)	0.35 (0.04)	$<2.0 \times 10^{-16}$ (A)	18.9
<b>Global Cognition</b>												
<i>rs3746319</i> (ZNF224)	11 (0.03)	AA	-0.759 (1.15)	113 (0.27)	AG	-1.19 (1.27)	289 (0.70)	GG	-0.79 (1.15)	-0.39 (0.12)	0.002 (D)	2.1
<i>rs8192708</i> (PCK1)	8 (0.02)	GG	-0.413 (0.670)	81 (0.20)	GA	-0.550 (0.947)	324 (0.78)	AA	-1.00 (1.24)	0.49 (0.13)	$1.02 \times 10^{-4}$ (A)	3.4
<i>APOE ε2</i>	2 (0.01)	-	-0.235 (1.11)	62 (0.15)	-	-0.788 (1.21)	339 (0.84)	-	-0.916 (1.19)	0.23 (0.15)	0.12 (A)	0.36
<i>APOE ε4</i>	9 (0.02)	-	-1.27 (1.50)	109 (0.27)	-	-1.41 (1.28)	285 (0.71)	-	-0.684 (1.09)	-0.69 (0.11)	$4.37 \times 10^{-10}$ (A)	9

<sup>1</sup> Mean quantitative trait outcome measure is reported, square root transformed for global pathology.

<sup>2</sup> Associations were tested with additive (A), dominant (D), or recessive models to identify the best fit.

doi:10.1371/journal.pone.0011244.t006

by 44% (Beta = -0.22,  $p = 0.05$ ). In our analyses of the neuropathologic subtypes, we found that the association of *ZNF224* on AD pathology appeared to be due to a predominant effect on neurofibrillary tangles. Indeed, when we substituted a term for neurofibrillary tangles instead of the global pathology variable in our regression model, the effect of the *ZNF224* variant on global cognition was reduced by 64%, and was no longer significant (Beta = -0.14,  $p = 0.21$ ), whereas tangles showed a robust association with cognitive impairment ( $p < 2 \times 10^{-16}$ ). These results are consistent with a sequence of events whereby an effect on the formation of neurofibrillary tangles accounts for the association of the *ZNF224* allele with cognitive function.

In contrast to *ZNF224*, the *PCK1* locus showed a relatively selective association with global cognition, but not with global pathology. For this locus, an additive model of inheritance was a best fit for our data, and *PCK1* explained 3.4% (Beta = 0.49,  $p = 1.02 \times 10^{-4}$ ) of the variance in global cognition proximate to death. We next used multiple linear regression to test whether the

association of this SNP on cognitive impairment is predominantly independent of AD pathology. Indeed, after inclusion of a model term for global AD pathology, the *PCK1* SNP (*rs8192708*) remained associated with global cognition (Beta = 0.37,  $p = 0.002$ ), despite the strong, independent association between pathology and cognition (Beta = -1.22,  $p < 2 \times 10^{-16}$ ; Table 7). Given our finding of an association with neuritic plaques, we substituted a model term for neuritic pathology instead of the global pathology variable; however, despite a modest reduction in the effect size, the relation between *PCK1* and cognitive impairment remained significant (Beta = 0.31,  $p = 0.005$ ; Table 7). Besides AD-related pathology, Lewy bodies and infarcts are the two additional brain pathologies most commonly seen in association with age-related cognitive decline [39]. We therefore investigated whether the association of *PCK1* on global cognition might be mediated by either Lewy bodies or infarcts, by including relevant terms into our regression model (Table 8). Again, the association between *PCK1* and global cognition remained

**Table 7.** Distinct pathways of *ZNF224* and *PCK1* association with cognition.

	Model 1 <sup>1</sup>		Model 2 <sup>2</sup>		Model 3 <sup>3</sup>			
	Beta (SE)	p	Beta (SE)	p	Beta (SE)	p		
<i>rs3746319</i> (ZNF224)	-0.39 (0.12)	0.002	-0.22 (0.11)	0.05	-0.14 (0.11)	0.213		
Pathology Measure	-	-	Global Pathology	-1.27 (0.13)	$<2 \times 10^{-16}$	Neurofibrillary Tangles	-1.40 (0.12)	$<2 \times 10^{-16}$
<i>rs8192708</i> (PCK1)	0.50 (0.120)	$1.02 \times 10^{-4}$	0.37 (0.12)	0.002	0.31 (0.11)	0.005		
Pathology Measure	-	-	Global Pathology	-1.22 (0.13)	$<2 \times 10^{-16}$	Neuritic Plaques	-0.96 (0.10)	$<2 \times 10^{-16}$

<sup>1</sup> Core regression model includes terms for age at death, gender, and years of education.

<sup>2</sup> Model additionally includes a term for the global AD pathology measure.

<sup>3</sup> Model additionally includes a term for either neurofibrillary tangles or neuritic plaques.

doi:10.1371/journal.pone.0011244.t007



**Table 8.** *PCK1* association with cognition is largely independent of AD pathology, infarcts, and Lewy bodies.

	Model 4 <sup>1</sup>	
	Beta (SE)	P
<i>rs8192708</i> ( <i>PCK1</i> )	0.39 (0.11)	$2.04 \times 10^{-4}$
AD Pathology	-1.24 (0.12)	$<2 \times 10^{-16}$
Infarcts	-0.39 (0.10)	$1.75 \times 10^{-4}$
Lewy bodies	-0.52 (0.11)	$3.84 \times 10^{-5}$

<sup>1</sup> Model also includes terms for age, gender, and education.  
doi:10.1371/journal.pone.0011244.t008

significant (Beta = 0.39,  $p = 2.04 \times 10^{-4}$ ), and *PCK1* continued to explain 3% of the residual variance in global cognition in our cohort after adjusting for the three most common brain pathologies associated with dementia. Common variation at the *PCK1* locus has also been associated with type 2 diabetes in a number of independent studies [40–42]. Since diabetes has also been implicated as a risk factor for age-related cognitive decline [31], we examined whether diabetes mediates the association of *PCK1* with cognitive impairment. However, adjusting for diabetes diagnosis in our linear regression model did not significantly attenuate the association of *PCK1* with global cognition (Beta = 0.370,  $p = 0.001$ ).

## Discussion

In this study, by genotyping a panel of loci within two cohorts of subjects with detailed cognitive and neuropathological characterization, we evaluate intermediate phenotypes as a tool for the functional dissection of candidate AD susceptibility loci. Using the extensively validated *APOE* locus, we previously demonstrated that intermediate traits enhance statistical power to detect associations, even in studies of modest sample size. Here, using the same strategy, we present evidence supporting the possible role of two additional loci in influencing age-related cognitive decline and AD neuropathology. Specifically, the *ZNF224* locus is associated with a quantitative measure of global AD pathology, and both *ZNF224* and *PCK1* are associated with a summary measure of global cognition proximate to death. Using separate quantitative traits for each of the predominant AD pathological features, we document associations between *GALP* and *PCK1* and diffuse and neuritic plaque pathology, respectively, whereas *ZNF224* showed a relatively selective association with neurofibrillary tangle pathology. Finally, in a series of statistical mediation analyses, we tested hypotheses about the causal chain of events linking genetic variation in the *ZNF224* and *PCK1* loci with cognitive decline, with strikingly different outcomes. In the case of *ZNF224*, we find that AD pathology, and more specifically, neurofibrillary tangles mediate an association with cognitive impairment. In contrast, we find that the association between *PCK1* and cognition is largely independent of not only AD pathology, but also Lewy bodies, and infarcts, which together comprise the three most common known brain pathologies associated with dementia [39,43].

Both *ZNF224* and *PCK1* were initially implicated by AD GWA studies; however, neither locus has yet been consistently replicated in subsequent genetic studies, and little is known about their potential mechanism of action in disease pathogenesis. The *ZNF224* locus encodes a Kruppel-associated box-containing zinc-finger protein that is widely expressed, including in the adult brain, and likely functions as a transcriptional repressor [44,45].

The SNP evaluated in this study, *rs3746319*, encodes a missense mutation causing a Lys to Glu change at position 640, which falls near the C-terminus within one of 19 zinc-finger repeat motifs. However, we do not yet know enough about *ZNF224* protein structure and function to speculate further on how this variant might promote neurofibrillary tangle formation and subsequent cognitive impairment, and further investigation will be required to determine if *rs3746319* is the causal variant and whether *ZNF224* is indeed the causal gene. The *PCK1* gene encodes phosphoenolpyruvate carboxykinase 1, which catalyzes the rate-limiting step of gluconeogenesis [46]. The SNP genotyped in our study, *rs8192708*, is also a missense mutation, causing an Ile to Val change at position 267; however, the functional consequences of this change, if any, are not known. *PCK1* variants have also been suggested to be associated with diabetes [40–42], and independently, diabetes has been identified as a risk factor for the development of dementia [31]. In our mediation analysis, adjusting for the effect of diabetes diagnosis did not account for the association of *PCK1* and cognitive impairment; however, it is possible that an appropriate intermediate phenotype, such as direct measurements of blood glucose or hemoglobin A1c, might allow detection of mediation. In another study performed in the same cohort, a relation was found between diabetes and infarcts [47]; however, we were also unable to mediate the *PCK1* association by including a model term for cerebral infarctions. Our finding that the *PCK1* association with cognitive decline is not explained by AD pathology, Lewy bodies, or infarcts suggests that this locus might influence additional, unmeasured pathologies. For example, whereas our analyses adjusted for macroscopic infarcts, *PCK1* may instead primarily influence microscopic forms of cerebrovascular injury. Further, while our intermediate pathologic phenotype accounts for amyloid plaques and neurofibrillary tangles, it does not capture levels of soluble, but potentially still neurotoxic, forms of amyloid or tau pathology [48,49]. Alternatively, variation at *PCK1* might influence one or multiple steps in the cascade of events predicted to occur downstream of amyloid, tangles and other pathologies, such as synapse loss, inflammation, and/or cell death pathways.

Unexpectedly, the variants in *ZNF224* and *PCK1* show opposite directions of allelic effects for association with AD intermediate phenotypes in our cohort compared to their association with AD diagnosis in the initial GWA studies. In other words, the alleles associated with increased AD risk in the initial reports (*rs3746319*<sup>G</sup> and *rs8192708*<sup>C</sup>) are actually protective against cognitive decline in our cohort. Importantly, this discrepancy is not accounted for by our use of intermediate phenotypes, as the *ZNF224* and *PCK1* SNPs show consistent direction of effect on AD diagnosis in our study population (Table 3). Such “flip-flop” associations have been reported with increasing frequency as GWA scans are completed for many common diseases, and replication efforts are subsequently undertaken [50]. Indeed, in the case of *PCK1*, two prior replication studies found evidence that the major allele, *rs8192708*<sup>A</sup>, may increase risk for dementia, consistent with our results suggesting an association between this allele and both cognitive decline and AD [4,38].

The interpretation of reversals in the direction of variant associations between different study cohorts remains controversial [50]. The most common explanation for such observations are that they are in fact spurious and representative of chance fluctuations around the null hypothesis. However, in our study, the strongly suggestive statistical evidence for the associations between *ZNF224* and *PCK1* with AD intermediate phenotypes makes their arising by chance less likely; and additionally, the reversals of allelic effect are seen with both loci in our analysis. Instead, we

propose that differences in subject ascertainment and recruitment are more likely to be responsible for our observations. The Religious Orders Study (ROS) and Rush Memory and Aging Project (MAP), from which our study cohort is based, are prospective, longitudinal studies in which subjects from the community are recruited non-demented at baseline (mean age = 75 and 79, for ROS and MAP respectively). All cases of clinical AD are therefore incident within our cohort. In contrast, similar to nearly all AD GWA studies performed to date, the initial reports of association with the *PCK1* and *ζNF224* loci come from AD cases recruited from a neurology clinic population with prevalent dementia. In addition, whereas subjects in our study were recruited at approximately similar ages to the GWA cohorts, they were significantly older at the time of last clinical evaluation and autopsy (mean age of death = 87). Studies with different designs (cross-sectional vs. prospective) and varying methods of subject ascertainment can generate contradictory epidemiological findings, for example due to survival bias. If an AD risk allele is associated with earlier age of dementia onset; it might be under-represented in the prospective cohort, which requires subjects to be non-demented at enrollment; and therefore, might subsequently appear to be a protective allele. “Flip-flop” associations might additionally arise from variation in linkage disequilibrium structure in the genomic region of interest between the cohorts in different studies. In fact, both the *ζNF224* and *PCK1* SNPs fall under modest recombination peaks, based on HapMap data [51]. Although both our study and the GWA analyses were conducted in subjects of European ancestry, it remains possible that sampling variation between two populations of similar ethnicity might lead to the association reversal that we have observed, as recombination could distribute our tag SNP onto haplotypes that are different from that harboring the causal variant [50]. Ultimately, further analysis of both SNPs and fine mapping of each locus in

larger study samples will be required to validate both *PCK1* and *ζNF224* as AD susceptibility loci, and resolve which allele may increase risk for disease.

Of the thirty-four SNPs evaluated in our study, both of the loci that we found to be associated with AD intermediate phenotypes were initially identified by GWA studies, suggesting the power of this unbiased approach to identify genes that might be overlooked by prevailing hypotheses of disease biology. Our study was initiated prior to the recent report of two large AD case/control GWA studies which independently identified three new susceptibility genes, *CLU*, *CRI*, and *PICALM* [15,16]. In a parallel effort, we recently found that *CRI* is associated with age-related cognitive decline in our study cohorts; and further, that this association was mediated by an effect on amyloid pathology (Chibnik et al., submitted). The power of a GWA study design and the types of genes one expects to discover are tightly linked to the selected phenotypic outcome. To the extent possible, the chosen outcome measure should be closely matched to the underlying biology responsible for the heritable trait variation of interest. In autopsy cohort studies of aged individuals in the community setting, most subjects with probable AD demonstrate multiple brain pathologies [52]. Based on our results, we believe that intermediate pathological and cognitive traits have great promise to enhance gene discovery and for functional characterization of loci that emerge from current AD GWA studies.

## Author Contributions

Conceived and designed the experiments: JMS JAS DB PLDJ. Performed the experiments: JMS CA. Analyzed the data: JMS LBC. Contributed reagents/materials/analysis tools: CA JAS DB. Wrote the paper: JMS LBC JAS DB PLDJ.

## References

- Ertekin-Taner N (2007) Genetics of Alzheimer's disease: a centennial review. *Neurol Clin* 25: 611–667.
- Bertram L, McQueen M, Mullin K, Blacker D, Tanzi R (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 39: 17–23.
- Cousin E, Macé S, Rocher C, Dib C, Muzard G, et al. (2009) No replication of genetic association between candidate polymorphisms and Alzheimer's disease. *Neurobiology of Aging*, in press.
- Feulner T, Laws S, Friedrich P, Wagenpfeil S, Wurst S, et al. (2009) Examination of the current top candidate genes for AD in a genome-wide association study. *Mol Psychiatry*, in press.
- Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, et al. (2008) A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC medical genomics* 1: 44.
- Beccham GW, Martin ER, Li Y-J, Slifer MA, Gilbert JR, et al. (2009) Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. *Am J Hum Genet* 84: 35–43.
- Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, et al. (2008) Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet* 83: 623–632.
- Bertram L, Tanzi RE (2009) Genome-wide association studies in Alzheimer's disease. *Human Molecular Genetics* 18: R137–145.
- Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, et al. (2009) Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet* 41: 192–198.
- Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, et al. (2007) Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum Mol Genet* 16: 865–873.
- Li H, Wetten S, Li L, St, Jean PL, Upmanyu R, et al. (2008) Candidate Single-Nucleotide Polymorphisms From a Genomewide Association Study of Alzheimer Disease. *Archives of Neurology* 65: 45–53.
- Li Y, Grupe A, Rowland C, Nowotny P, Kauwe J, et al. (2006) DAPK1 variants are associated with Alzheimer's disease and allele-specific expression. *Hum Mol Genet* 15: 2560–2568.
- Liu F, Arias-Vásquez A, Sleegers K, Aulchenko YS, Kayser M, et al. (2007) A genome-wide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. *Am J Hum Genet* 81: 17–31.
- Reiman E, Webster J, Myers A, Hardy J, Dunckley T, et al. (2007) GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 54: 713–720.
- Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, et al. (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 41: 1088–1093.
- Lambert J-C, Heath S, Even G, Campion D, Sleegers K, et al. (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 41: 1094–1099.
- Bennett D, Schneider J, Arvanitakis Z, Kelly J, Aggarwal N, et al. (2006) Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* 66: 1837–1844.
- McQueen MB, Bertram L, Lange C, Becker KD, Albert MS, et al. (2007) Exploring candidate gene associations with neuropsychological performance. *Am J Med Genet* 144B: 987–991.
- Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, et al. (2009) Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. *PLoS ONE* 4: e6501.
- Seshadri S, DeStefano A, Au R, Massaro J, Beiser A, et al. (2007) Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med Genet* 8 Suppl 1: S15.
- Papassotiropoulos A, Streffer JR, Tzolaki M, Schmid S, Thal D, et al. (2003) Increased brain beta-amyloid load, phosphorylated tau, and risk of Alzheimer disease associated with an intronic CYP46 polymorphism. *Archives of Neurology* 60: 29–35.
- Peskind E, Li G, Shofer J, Quinn J, Kaye J, et al. (2006) Age and apolipoprotein E\*4 allele effects on cerebrospinal fluid beta-amyloid 42 in adults with normal cognition. *Arch Neurol* 63: 936–939.
- Bennett DA, De Jager PL, Leurgans SE, Schneider JA (2009) Neuropathologic intermediate phenotypes enhance association to Alzheimer susceptibility alleles. *Neurology* 72: 1495–1503.
- Bennett D, Schneider J, Wilson R, Bienias J, Berry-Kravis E, et al. (2005) Amyloid mediates the association of apolipoprotein E e4 allele to cognitive function in older people. *J Neurol Neurosurg Psychiatry* 76: 1194–1199.
- Mortimer JA, Snowdon DA, Markesbery WR (2009) The effect of APOE-epsilon4 on dementia is mediated by Alzheimer neuropathology. *Alzheimer Dis Assoc Disord* 23: 152–157.

26. Bertram L, Tanzi RE (2008) Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nat Rev Neurosci* 9: 768–778.
27. McKhann G, Drachman D, Folstein M, Katzman R, Price D, et al. (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34: 939–944.
28. Bennett DA, Schneider JA, Aggarwal NT, Arvanitakis Z, Shah RC, et al. (2006) Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology* 27: 169–176.
29. Bennett DA, Wilson RS, Schneider JA, Evans DA, Beckett LA, et al. (2002) Natural history of mild cognitive impairment in older persons. *Neurology* 59: 198–205.
30. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12: 189–198.
31. Arvanitakis Z, Wilson RS, Bienias JL, Evans DA, Bennett DA (2004) Diabetes mellitus and risk of Alzheimer disease and decline in cognitive function. *Archives of Neurology* 61: 661–666.
32. The National Institute on Aging and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease (1997) Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. *Neurobiol Aging* 18: S1–2.
33. Braak H, Braak E (1991) Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 82: 239–259.
34. Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, et al. (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* 41: 479–486.
35. Bennett DA, Schneider JA, Tang Y, Arnold SE, Wilson RS (2006) The effect of social networks on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. *Lancet Neurol* 5: 406–412.
36. Bennett DA, Wilson RS, Schneider JA, Evans DA, Aggarwal NT, et al. (2003) Apolipoprotein E epsilon4 allele, AD pathology, and the clinical expression of Alzheimer's disease. *Neurology* 60: 246–252.
37. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.
38. Figgins JA, Minster RL, Demirci FY, Dekosky ST, Kamboh MI (2009) Association studies of 22 candidate SNPs with late-onset Alzheimer's disease. *Am J Med Genet B Neuropsychiatr Genet* 150B: 520–526.
39. Sonnen JA, Larson EB, Crane PK, Haneuse S, Li G, et al. (2007) Pathological correlates of dementia in a longitudinal, population-based sample of aging. *Ann Neurol* 62: 406–413.
40. Cao H, van der Veer E, Ban MR, Hanley AJG, Zinman B, et al. (2004) Promoter polymorphism in PCK1 (phosphoenolpyruvate carboxykinase gene) associated with type 2 diabetes mellitus. *J Clin Endocrinol Metab* 89: 898–903.
41. Hamilton G, Proitsi P, Jehu L, Morgan A, Williams J, et al. (2007) Candidate gene association study of insulin signaling genes and Alzheimer's disease: Evidence for SOS2, PCK1, and PPAR $\gamma$ s susceptibility loci. *Am J Med Genet* 144B: 508–516.
42. Willer CJ, Bonnycastle LL, Conneely KN, Duren WL, Jackson AU, et al. (2007) Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. *Diabetes* 56: 256–264.
43. Schneider J, Arvanitakis Z, Bang W, Bennett D (2007) Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology* 69: 2197–2204.
44. Medugno L, Florio F, Cesaro E, Grosso M, Lupo A, et al. (2007) Differential expression and cellular localization of ZNF224 and ZNF255, two isoforms of the Krüppel-like zinc-finger protein family. *Gene* 403: 125–131.
45. Medugno L, Florio F, De Cegli R, Grosso M, Lupo A, et al. (2005) The Krüppel-like zinc-finger protein ZNF224 represses aldolase A gene transcription by interacting with the KAP-1 co-repressor protein. *Gene* 359: 35–43.
46. Yang J, Kalhan SC, Hanson RW (2009) What is the metabolic role of phosphoenolpyruvate carboxykinase? *J Biol Chem* 284: 27025–27029.
47. Arvanitakis Z, Schneider JA, Wilson RS, Li Y, Arnold SE, et al. (2006) Diabetes is related to cerebral infarction but not to AD pathology in older persons. *Neurology* 67: 1960–1965.
48. McLean CA, Cherny RA, Fraser FW, Fuller SJ, Smith MJ, et al. (1999) Soluble pool of Abeta amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Ann Neurol* 46: 860–866.
49. Santacruz K, Lewis J, Spire T, Paulson J, Kotilinek L, et al. (2005) Tau suppression in a neurodegenerative mouse model improves memory function. *Science* 309: 476–481.
50. Lin P-I, Vance JM, Pericak-Vance MA, Martin ER (2007) No gene is an island: the flip-flop phenomenon. *Am J Hum Genet* 80: 531–538.
51. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851–861.
52. Schneider JA, Arvanitakis Z, Leurgans SE, Bennett DA (2009) The neuropathology of probable Alzheimer disease and mild cognitive impairment. *Ann Neurol* 66: 200–208.