Genome-wide pleiotropy analysis of neuropathological traits related to Alzheimer’s disease

Jaeyoon Chung1,2, Xiaoling Zhang2, Mariet Allen3, Xue Wang4, Yiyi Ma2, Gary Beecham5, Thomas J. Monte6, Steven G. Younkin3, Dennis W. Dickson3, Todd E. Golde7, Nathan D. Price8, Nilüfer Ertekin-Taner3,9, Kathryn L. Lunetta10, Jesse Mez11, Alzheimer’s Disease Genetics Consortium, Richard Mayeux12, Jonathan L. Haines13, Margaret A. Pericak-Vance5, Gerard Schellenberg14, Gyungah R. Jun2,15 and Lindsay A. Farrer1,2,10,11,16,17*

Abstract

Background: Simultaneous consideration of two neuropathological traits related to Alzheimer’s disease (AD) has not been attempted in a genome-wide association study.

Methods: We conducted genome-wide pleiotropy analyses using association summary statistics from the Beecham et al. study (PLoS Genet 10:e1004606, 2014) for AD-related neuropathological traits, including neuritic plaque (NP), neurofibrillary tangle (NFT), and cerebral amyloid angiopathy (CAA). Significant findings were further examined by expression quantitative trait locus and differentially expressed gene analyses in AD vs. control brains using gene expression data.

Results: Genome-wide significant pleiotropic associations were observed for the joint model of NP and NFT (NP + NFT) with the single-nucleotide polymorphism (SNP) rs34487851 upstream of C2orf40 (alias ECRG4, $P = 2.4 \times 10^{-8}$) and for the joint model of NFT and CAA (NFT + CAA) with the HDAC9 SNP rs79524815 ($P = 1.1 \times 10^{-8}$). Gene-based testing revealed study-wide significant associations ($P \leq 2.0 \times 10^{-6}$) for the NFT + CAA outcome with adjacent genes TRAPPCL2, TRAPPCL2-AS1, and ADI1. Risk alleles of proxy SNPs for rs79524815 were associated with significantly lower expression of HDAC9 in the brain ($P = 3.0 \times 10^{-3}$), and HDAC9 was significantly downregulated in subjects with AD compared with control subjects in the prefrontal ($P = 7.9 \times 10^{-5}$) and visual ($P = 5.6 \times 10^{-4}$) cortices.

Conclusions: Our findings suggest that pleiotropy analysis is a useful approach to identifying novel genetic associations with complex diseases and their endophenotypes. Functional studies are needed to determine whether ECRG4 or HDAC9 is plausible as a therapeutic target.

Keywords: Alzheimer’s disease, Neuropathological traits, Genome-wide association study, Pleiotropy analysis, HDAC9, ECRG4

Background

Alzheimer’s disease (AD) is the most common type of dementia in persons aged 65 years and older [1, 2]. Pathologically, it is characterized primarily by the appearance of both neuritic plaques (NPs) containing oligomers of β-amyloid and neurofibrillary tangles (NFTs), accompanied by a progressive loss of neurons in the brain [3, 4]. Also, cerebral amyloid angiopathy (CAA), which is caused by aggregates of β-amyloid in walls of blood vessels in the brain, is found in as many as 90% of autopsy-confirmed AD cases [5]. Previously, Beecham et al. identified multiple significant gene associations in a genome-wide association study (GWAS) for several AD-related neuropathological traits, including NP, NFT, and CAA measured in brains from subjects with pathologically confirmed AD cases and from control subjects with no...
evidence of neurological disease [6]. We hypothesized that additional novel associations could be identified in models allowing a genetic variant to influence more than one trait (i.e., pleiotropy). In this study, we performed genome-wide pleiotropy analyses of joint models of NP, NFT, and CAA using summary data from the previous study [6].

Methods
Study population, neuropathological trait selection, and data processing
We obtained summary statistics from univariate GWAS of NP, NFT, and CAA [6]. These results were derived from meta-analyses of 12 studies including 3598 subjects (3135 AD cases, 463 controls) of European ancestry. Neuropathological data for the entire sample were reviewed and harmonized by one neuropathologist for consistency across studies [6]. Although Beecham et al. [6] also evaluated Lewy body disease, hippocampal sclerosis, and vascular brain disease, we limited our present analyses to neuropathological outcomes most directly linked to AD and moderately correlated with each other (i.e., NP, NFT, and CAA). Uncorrelated traits are unlikely to show significant pleiotropic associations, and results from pleiotropy analysis will be similar to those from univariate models (i.e., single phenotype) if the traits are highly correlated. Details of subject recruitment, genotyping, genotype imputation, quality control procedures, population substructure analysis, and statistical methods for association analyses of individual traits were reported previously [6, 7]. Sample demography of the 3598 subjects with autopsied brains and genotypes (3135 cases and 463 controls) is described in Additional file 1: Table S1.

Univariate genome-wide association analyses
Results from the association tests by Beecham et al. [6] in each dataset for each neuropathologic trait with genotypes imputed using the 1000 Genomes Project reference panel (GRCh37 at December 2010) for a genome-wide set of single-nucleotide polymorphisms (SNPs) were obtained using ordinal logistic regression models including the first three principal components of ancestry as covariates to account for population substructure [6]. NP and NFT measures were analyzed in well-established ordinal rankings (NPs: none, sparse, moderate, and frequent by Consortium to Establish a Registry for Alzheimer’s Disease (“CERAD”) scoring [8]; NFT: none, transentorhinal, limbic, and isocortical by Braak and Braak staging [9]), and CAA was analyzed as a binary trait (present or absent). Full details of these analyses are reported elsewhere [6]. We used GWAS meta-analysis summary statistics (β and SE) of the three neuropathologic traits for 6.5 million imputed SNPs after omitting SNPs from studies if the minor allele frequency was ≤ 1%, imputation quality ($R^2$) was ≤ 0.4, or dosage variance was ≤ 0.02.

Genome-wide pleiotropy analyses
We conducted a genome-wide pleiotropy analysis for each pair of the three neuropathological traits using the O’Brien method [10, 11], which is implemented in an R library (“CUMP”) [12]. This method combines univariate test statistics (Z-scores from β and SE values) of all SNPs from separate GWASs for individual phenotypes to compute a test statistic that follows a multivariate normal distribution. The covariance matrix of the distribution was approximated by the sample covariance matrix of the test statistics of all SNPs. Under the null hypothesis, an SNP is not associated with any of the phenotypes. The alternative hypothesis is that an SNP is associated with at least one of the phenotypes. We defined a SNP as having a pleiotropic effect on two phenotypes when the $P$ value for the O’Brien test statistic from the joint model of association of two phenotypes ($P_{\text{joint}}$) with the SNP is at least one order of magnitude more significant than the $P$ values ($P_{\text{univariate}}$) for both phenotypes and the univariate $P$ values are at least nominally significant ($P_{\text{univariate}} < 0.05$). As a supplementary analysis, we also conducted a trivariate pleiotropy genome-wide analysis for the three neuropathological traits. The genome-wide significance (GWS) threshold for these analyses was set at $P < 5.0 \times 10^{-8}$.

Gene-based association
We performed genome-wide gene-based tests for each joint model using results from individual SNP tests. SNPs within 30 kb of the transcription start and end sites were included in each gene-based test. These analyses were carried out using the versatile gene-based test (“VEGAS”) method [13], which computes an empirical $P$ value through Monte Carlo simulations based on linkage disequilibrium patterns of the European ancestry population in the 1000 Genomes Project (GRCh37 released March 2012). The GWS level for the gene-based tests was set at $2.7 \times 10^{-6}$, which was calculated as the nominal significance level 0.05 divided by the total number of genes tested ($n = 18,500$).

Expression quantitative trait locus analysis
The association of SNP genotypes with gene-level expression (i.e., expression quantitative trait loci [eQTLs]) was evaluated using version 6 of the GTEx Portal database (http://www.gtexportal.org/; [14]) and data from the Mayo Clinic brain expression GWAS (eGWAS) (https://www.synapse.org/#!Synapse:syn3157249 or http://alos.med.upenn.edu/niagads; [15]). The GTEx Portal provides eQTL association summary statistics (β and $P$ values) across 43 different tissues from 175 subjects. The Mayo Clinic brain
eGWAS data were generated from the cerebellum (CER; 197 AD and 177 non-AD control subjects) and temporal cortex (TCX; 202 AD and 197 non-AD control subjects) regions. Gene expression measures for 24,526 probes were generated with the Illumina Whole Genome DASL array (Illumina, San Diego, CA, USA). SNP genotype data for the Mayo Clinic eGWAS were obtained from the Mayo Clinic late-onset AD GWAS [16]. AD cases were diagnosed as definite AD according to National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer’s Disease and Related Disorders Association criteria, whereas non-AD controls had other neuropathologies. For each brain region, association of gene expression and imputed SNP genotype (GRCh36) was evaluated using linear regression, including covariates for AD status, apolipoprotein E (APOE) ε4 dosage (0, 1, or 2), age at death, sex, plate, RNA integrity number (RIN), and adjusted RIN (RIN – RINmean²). Analyses were also conducted for AD cases and controls separately.

Differential gene expression analysis

Differential gene expression (DGE) analysis was performed using publicly available brain whole-transcriptome RNA-sequencing (RNA-Seq) data [17] and microarray data (Gene Expression Omnibus accession number [GEO:GSE44772] [18]). The RNA-Seq data include DGE summary statistics for the CER and TCX derived from 86 subjects with AD and 80 control subjects (https://www.synapse.org). Following a quality control step, 80 AD and 76 control brains were analyzed. All subjects underwent RNA-Seq using the Illumina HiSeq 2000 sequencing system (101 bp, paired-end sequencing) at the Mayo Clinic Genomic Core Facility. All AD and some of the control brains were from the Mayo Clinic Brain Bank, whereas other control brains were from the Banner Sun Health Research Institute (Sun City, AZ, USA). Following quality control, raw read counts normalized according to conditional quantile normalization (CQN) employing the Bioconductor package were used in the analyses. For DGE comparing AD with controls, multivariable linear regression analyses were conducted in R, using CQN normalized gene expression measures and including age at death, sex, RIN, brain tissue source, and flow cortex as biological and technical covariates. To account for any CNS cell-population changes that occur as a consequence of disease pathology, we also included cell-specific sequence of disease pathology, we also included cell-specific gene levels as covariates, using the expression levels for the five central nervous system (CNS)-specific genes as follows: ENO2 for neurons [ENCODE:ENSG00000116764], GFAP for astrocytes [ENCODE:ENSG00000131095], CD68 for microglia [ENCODE:ENSG00000129226], OLIG2 for oligodendrocytes [ENCODE:ENSG00000205927], and CD34 for endothelial cells [ENCODE:ENSG00000174059].

The microarray gene expression data were generated from autopsied brains collected from dorsolateral prefrontal cortex (DLPFC), visual cortex (VCX), and CER regions of 129 AD patients and 101 control subjects. Samples were profiled on a custom-made Agilent 44K array (Agilent Technologies, Santa Clara, CA, USA) containing 40,638 human genes. Gene expression data were normalized using Rosetta Resolver gene expression analysis software as previously described [18]. The association between expression of each gene (outcome) and AD status (predictor) was tested using linear regression adjusting for RIN, postmortem interval, batch, preservation method, pH in tissues, age, sex, and the five cell-type markers.

Results

NP, NFT, and CAA were moderately correlated (NP-NFT, r = 0.68; NP-CAA, r = 0.56; NFT-CAA, r = 0.40; P < 2.2 × 10⁻¹⁶ for each pair of traits), indicating a potential for discovery of novel associations in pleiotropy analysis.

Bivariate GWAS results

There was no inflation in P values for the GWAS of the three neuropathological traits analyzed individually (genomic control parameter, λ = 1.00, 1.01, and 0.96 for NP, NFT, and CAA, respectively) or as joint outcomes (Additional file 1: Figure S1). Results of the pleiotropy GWAS are shown in Additional file 1: Figure S2. As reported previously, with the exception of APOE, only 15 of 25 previously known AD loci attained at least a nominal association with NP, NFT, or CAA [6]. Three of the previously established AD loci—BIRN1, HLA, and PICALM—were moderately associated (P < 10⁻⁴) in the pleiotropy analysis for NP and NFT at a significance level of at least one order of magnitude smaller compared with the results from univariate analyses (Additional file 1: Table S2). Two novel GWS associations were detected in the pleiotropy analyses (Table 1, Fig. 1). rs34487851, an SNP located approximately 40 kb upstream of C2orf80, was associated with the joint model of NP and NFT (Pjoint = 2.0 × 10⁻⁶). An intronic SNP in HDAC9, rs79524815, was associated with the joint model of NFT and CAA (Pjoint = 1.1 × 10⁻⁸). The major allele A of rs34487851 and the minor allele G of rs79524815 are associated with increased NP and NFT and with increased NFT and the presence of CAA, respectively. Both of these findings were at least one order of magnitude more significant than for the univariate traits (Table 1) and were supported by evidence from multiple SNPs at those locations (Fig. 1, Additional file 1: Table S3).

Bivariate gene-based pleiotropy analysis results

Three contiguous novel genes on chromosome 2p25.3 (TRAPPC12, TRAPPC12-AS1, and ADI1) were associated with the joint model of NFT and CAA at a gene-wide significant level (P ≤ 2.0 × 10⁻⁶) (Table 2 and Additional file 1: Figure S3). Of note, one SNP in this region (rs35067331 in TRAPPC12) was associated with the NFT-
Table 1: Genome-wide significant association ($P < 5.0 \times 10^{-8}$) of novel genes in the genome-wide pleiotropy analyses (joint models) of the three neuropathological traits neuritic plaque, neurofibrillary tangles, and cerebral amyloid angiopathy.

| Chromosome | SNP | Gene | EA | RA | EAF | Univariate models | Joint models |
|------------|-----|------|----|----|-----|-------------------|--------------|
|            |     |      |    |    |     | AD status | NP | NFT | CAA | NP + NFT | NP + CAA | NFT + CAA |
|            |     |      |    |    |     | β value (SE) | P value | β value (SE) | P value | β value (SE) | P value | β value (SE) | P value |
| 2          | rs34487851 | ECRG4 | G | A | 0.27 | 0.42 (0.09) | 5.8 × 10^{-6} | -0.3 (0.06) | -0.25 (0.06) | 4.5 × 10^{-6} | -0.14 (0.08) | 0.06 |
|            |     |      |    |    |     | -2.0 × 10^{-8} | - | 2.5 × 10^{-6} | - | 2.1 × 10^{-5} | - |
| 7          | rs79524815 | HDAC9 | G | T | 0.03 | 0.69 (0.31) | 0.03 | 0.03 | 0.79 (0.19) | 2.3 × 10^{-5} | 1.16 (0.26) | 9.1 × 10^{-6} |
|            |     |      |    |    |     | + | 1.3 × 10^{-4} | + | 3.3 × 10^{-6} | + | 1.1 × 10^{-8} |

Abbreviations: EA Effect allele, RA Reference allele, EAF Effect allele frequency, SNP Single-nucleotide polymorphism, AD Alzheimer's disease, NP Neuritic plaque, NFT Neurofibrillary tangles, CAA Cerebral amyloid angiopathy.
*Also known as C2orf40.
CAA outcome at nearly the GWS level \((P_{\text{joint}} = 5.8 \times 10^{-8})\) (Additional file 1: Table S4).

**Trivariate GWAS and gene-based pleiotropy analysis results**

We conducted trivariate GWAS and gene-based association analyses to identify genetic factors common to NP, NFT, and CAA. There was no evidence for genomic inflation \((\lambda = 0.99)\) in the results from the trivariate model (Additional file 1: Figure S4). GWS association was observed only for \(APOE\) isoform SNP rs429358 \((P = 2.1 \times 10^{-47})\), whereas associations at \(C2orf40, HDAC9\), and \(TRAPPC12\) were attenuated (Additional file 1: Table S5).
**Table 2** Gene-wide significant results ($P < 2.7 \times 10^{-5}$) from gene-based tests of pleiotropy single-nucleotide polymorphism association results

| Chromosome | Start | End  | Gene            | Univariate gene-based tests | Pleiotropy gene-based tests |
|------------|-------|------|-----------------|----------------------------|-----------------------------|
|            |       |      |                 | NP  | NFT  | CAA | NP + NFT | NP + CAA | NFT + CAA |
| 2          | 3,383,446 | 3,483,342 | TRAPPC12       | 0.09 | 4.0 × 10^{-5} | 0.5 | 0.01 | 0.07 | 2.0 × 10^{-5} |
| 2          | 3,485,013 | 3,486,180 | TRAPPC12-AS1    | 0.002 | 3.9 × 10^{-5} | 5.0 × 10^{-5} | 1.6 × 10^{-5} | 2.1 × 10^{-5} | <1.0 × 10^{-6} |
| 2          | 3,501,690 | 3,523,350 | ADI1           | 0.003 | 1.6 × 10^{-5} | 7.0 × 10^{-4} | 3.2 × 10^{-5} | 4.0 × 10^{-6} | <1.0 × 10^{-6} |

**Abbreviations:** NP Neuritic plaque, NFT Neurofibrillary tangles, CAA Cerebral amyloid angiopathy

Gene-based $P$ values were computed through 1 million permutations, so the smallest $P$ value is $1.0 \times 10^{-6}$

**eQTL analysis**

We performed eQTL association analysis to examine whether the expression levels of the five GWS significant genes identified in the pleiotropy association tests differed between carriers and noncarriers of the risk alleles from those loci. Because information about the two GWS SNPs was not available in the GTEx Portal database or in the Mayo Clinic brain eGWAS, we analyzed proxy SNPs that are in high linkage disequilibrium ($LD; D’ \geq 80$) with the GWS SNPs. According to GTEx, rs34487785 proxy SNP rs1232803 is a cis-acting eQTL, and the major allele $A$, which is associated with higher NP and NFT, is also significantly associated with decreased expression of $C2orf40$ in several tissues, including the esophagus ($P = 3.5 \times 10^{-5}$), transverse colon ($P = 4.7 \times 10^{-4}$), and tibial artery ($P = 1.7 \times 10^{-3}$), but not in any brain regions. In the Mayo Clinic brain eGWAS, proxy SNPs for $C2orf40$ were not cis-acting eQTLs for $C2orf40$. In GTEx, proxy SNPs for rs79524815 were not associated with the expression of $HDAC9$. However, in the brain eGWAS, the minor alleles of proxy SNPs for rs79524815, which are associated with higher NFT and CAA, were significantly associated with lower $HDAC9$ levels in the CER (probe ID: ILMN_1803563; best eQTL, rs4721721; $P = 0.03$) but not in the TCX (Additional file 1: Table S3). According to GTEx, rs35067331 is a cis-acting eQTL, and its major allele $C$, which is associated with higher NFT and CAA, is significantly associated with increased expression of $TRAPPC12-AS1$ in several brain regions (best $P = 2.1 \times 10^{-7}$ in cortex) and $ADI1$ ($P = 0.03$) in the caudate nucleus, but not with differential expression of $TRAPPC12$ in any brain regions (Additional file 1: Table S4). In the Mayo Clinic brain eGWAS data, rs35067331 and its proxy SNPs were not cis-acting eQTLs for $ADI1$ or $TRAPPC12$. Unfortunately, information about $TRAPPC12-AS1$ was unavailable in the brain eGWAS.

**Differential gene expression analysis**

We investigated whether the expression levels of $C2orf40$, $HDAC9$, and $TRAPPC12/TRAPPC12-AS1/ADI1$ differed in AD brains compared with non-AD control brains in the publicly available RNA-Seq and microarray datasets (Table 3 and Fig. 2). There were no significant differences in $C2orf40$ expression between subjects with AD and control subjects in the TCX or CER in the Mayo Clinic RNA-Seq DGE profiling. However, $C2orf40$ was significantly downregulated in subjects with AD compared with control subjects in the CER ($P = 1.6 \times 10^{-3}$), DLPFC ($P = 0.04$), and VCX ($P = 2.7 \times 10^{-3}$) in the microarray brain expression data. $HDAC9$ was significantly downregulated in subjects with AD compared with control subjects in several brain regions, including the TCX ($P = 1.5 \times 10^{-4}$) and CER ($P = 0.04$) in the RNA-Seq profiling data and in the DLPFC.

**Table 3** Results of differential gene expression analysis in brain

| Gene       | RNA-Seq | Microarray | DLPFC | VCX |
|------------|---------|------------|-------|-----|
|            | β value (SE) | P value | β value (SE) | P value | β value (SE) | P value | β value (SE) | P value |
| ECGR4      | 0.06 (0.20) | 0.77 | 0.19 (0.24) | 0.43 | −0.18 (0.05) | 1.6 × 10^{-3} | −0.12 (0.06) | 0.04 | −0.12 (0.04) | 2.7 × 10^{-3} |
| HDAC9      | −0.24 (0.12) | 0.04 | −0.31 (0.08) | 1.5 × 10^{-4} | −0.01 (0.02) | 0.77 | −0.09 (0.03) | 7.9 × 10^{-3} | −0.06 (0.02) | 5.6 × 10^{-4} |
| TRAPPC12   | −0.05 (0.06) | 0.35 | −0.13 (0.05) | 0.01 | −0.09 (0.03) | 1.1 × 10^{-3} | −0.03 (0.02) | 0.09 | −0.08 (0.02) | 3.2 × 10^{-4} |
| TRAPPC12-AS1 | 0.22 (0.12) | 0.06 | 0.59 (0.18) | 1.3 × 10^{-3} | − | − | − | − | − |
| ADI1       | −0.10 (0.08) | 0.19 | −0.07 (0.08) | 0.36 | −0.10 (0.03) | 4.9 × 10^{-4} | −0.03 (0.03) | 0.31 | −0.01 (0.03) | 0.64 |

**Abbreviations:** CER Cerebellum, TCX Temporal cortex, DLPFC Dorosilateral prefrontal cortex, CER Cerebellum, TCX Temporal cortex, DLPFC Dorosilateral prefrontal cortex, VCX Visual cortex

Results were obtained from analyses of RNA-Seq data in the Synapse database (https://www.synapse.org) and microarray data in the Gene Expression Omnibus database (GEO:GSE44771). Negative β value indicates lower level of gene expression in AD cases compared with controls and vice versa. Results that remained significant after multiple test correction ($P = 0.05/22 = 2.27 \times 10^{-5}$) are highlighted in **bold**

*Also known as C2orf40*
Fig. 2 Box plots showing differential expression in microarray data [GEO:GSE44772] between AD cases and controls for C2orf40, HDAC9, TRAPPC12, and ADI1 in the cerebellum (CER; left column), dorsolateral prefrontal cortex (DLPFC; middle column), and visual cortex (VCX; right column). AD Alzheimer’s disease.
(P = 7.9 × 10^{-3}) and VCX (P = 5.6 × 10^{-4}) in the microarray expression data. ADI1 expression was downregulated in subjects with AD in the CER in the microarray data (P = 4.9 × 10^{-4}). The RNA-Seq DGE profiling indicated that TRAPPC12-AS1 expression was significantly increased in subjects with AD in the TCX (P = 1.3 × 10^{-3}). In contrast, expression of TRAPPC12 was significantly lower in subjects with AD than in control subjects in the TCX (P = 0.01) in the RNA-Seq and in the CER (P = 1.1 × 10^{-3}) and VCX (P = 3.2 × 10^{-4}) in the microarray data. Information about expression of TRAPPC12-AS1 was not available in the microarray data. The majority of these nominally significant findings survived correction for multiple testing (P < 2.27 × 10^{-3}).

To contextualize our findings, we evaluated DGE among AD cases and controls for 26 previously established AD genes (Additional file 1: Table S6) [19–21]. With notable exceptions of EPHA1 (P = 2.8 × 10^{-5}) and SLC24A4 (P = 7.0 × 10^{-5}) in CER in the RNA-Seq data and ABCA7 in DLPFC (P = 2.5 × 10^{-5}) and VCX (P = 4.9 × 10^{-5}) in the microarray data, none of the results for the other 23 genes were significant after correcting for 127 tests (P < 3.94 × 10^{-4}).

**Discussion**

A previous GWAS of neuropathologic traits including NP, NFT, and CAA identified GWS associations with APOE only [6]. Our pleiotropy analysis of all pairwise combinations of these traits identified GWS associations with APOE and three regions not previously reported with any neuropathologic traits or AD risk, including C2orf40 for the joint model comprising NP and NFT, as well as HDAC9 and TRAPPC12/TRAPPC12-AS1/ADI1 for the joint model comprising NFT and CAA. Our DGE study found that HDAC9 is significantly downregulated in several brain regions in subjects with AD compared with control subjects. Moreover, we observed that the G allele of HDAC9 SNP rs79524815 is associated with a higher level of the joint outcome of NFT and CAA, and proxy SNPs for rs79524815 (which are suggestively associated with the joint outcome of NFT and CAA) are associated with decreased HDAC9 expression in subjects with AD (Additional file 1: Table S3). The pleiotropy analysis also revealed that 4 (BIN1, HLA, PICALM, and APOE) of the 25 previously reported GWS AD risk loci [7, 19–21] were at least one order of magnitude more significantly associated with the joint model of NP + NFT than each of these traits analyzed separately, suggesting that these genes are involved in pathways leading to both plaques and tangles [22–25].

It is notable that pleiotropy analyses for the model including NP, NFT, and CAA did not yield any GWS findings. Moreover, the GWS associations identified in the bivariate models were attenuated in the trivariate model. These results suggest that the mechanisms or pathways underlying the bivariate associations probably do not encompass all three traits, and this conclusion may generalize genome-wide.

C2orf40, also known as esophageal cancer-related gene 4 (ECRG4), is a tumor suppressor gene [26] that encodes a peptide hormone that is involved in NFT formation in transgenic mice [27], senescence of precursor cells in the CNS during aging [28], and activation of microglia and peripheral mononuclear leukocytes [29]. We observed that ECRG4 expression was significantly lower in AD cases than in controls in several brain regions. Abnormal downregulation of C2orf40 was previously reported in brain injury [30] as well as in several cancer cell types [31, 32].

HDAC9 encodes a member of class IIa histone genes that deacetylate histones, thereby remodeling chromatin structure and controlling gene expression [33, 34] that has previously been linked to epigenetic mechanisms [35] and memory loss [36] in AD and also has been proposed as a possible therapeutic target [37–39]. GWS association of ischemic stroke with an HDAC9 variant was identified by GWAS [40]. Structural variants including deletions and copy number variants in HDAC9 have been identified in patients with schizophrenia and patients with autism [41, 42]. MEF2C, one of the well-established AD risk loci [21], stimulates HDAC9 expression, but HDAC9 suppresses MEF2C transcription, resulting in a negative feedback loop [43]. In a previously reported coexpression network study in AD and control brains, HDAC9 and MEF2C were clustered together in the top fourth module ranked by relevance to AD pathology, and expression of HDAC9 and MEF2C was inversely correlated with Braak stage (HDAC9, r = −0.71; MEF2C, r = −0.65) and frontal atrophy (HDAC9, r = −0.57, MEF2C, r = −0.51) [18]. These findings are consistent with our observation that HDAC9 expression is reduced in subjects with AD and in the subjects with HDAC9 SNP alleles associated with higher NFT and CAA. Decreased HDAC9 expression has also been linked to increased neuronal apoptosis [44, 45]. Collectively, findings from our and other studies indicate that MEF2C and HDAC9 may participate in a pathway leading to NFT formation and brain atrophy.

Gene-based analyses identified significant associations with three adjacent loci—TRAPPC12, TRAPPC12-AS1, and ADI1—in a gene-rich region near the end of the short arm of chromosome 2p. ADI1, encoding acireductone dioxygenase 1, is involved in methionine salvage...
and prostate cancer [46] and has no known relationship to AD. TRAPPC12 is a subunit of a trafficking protein particle complex that has a role in vesicle trafficking in endoplasmic reticulum (ER) to Golgi [47]. TRAPPC12-AS1 is an antisense (noncoding) RNA that contains a 1168 transcript from TRAPPC12. We previously established that regulation of vesicular trafficking in the ER to Golgi by several VPS10 receptor domain receptor genes, including SORLI, and by other genes encoding members of the retromer complex is an important pathway leading to AD [48–50]. Of the genes in this region, only TRAPPC12-AS1 showed a pattern of expression in subjects with AD and control subjects that is consistent with the effect direction of the TRAPPC12 rs35067331 allele’s influence on NFT and CAA. It should be noted that TRAPPC12 expression was significantly lower in AD cases than in control subjects in the TCX (P = 0.01) in the RNA-Seq data and the CER (P = 1.1 × 10⁻⁴) and VCX (P = 3.2 × 10⁻⁴) in the microarray data, which could be due to negative feedback by the antisense TRAPPC12-AS1 transcript [51].

Our study has several potential caveats. The GWS associations identified in the pleiotropy analysis have moderate supportive evidence for association from other SNPs under the association peaks, probably because of low LD with the top SNPs. The two GWS SNPs near C2orf40 and HDAC9 were not associated with AD risk in one of the largest GWASs for AD (rs34487851, P = 0.07; rs79524815, P = 0.73) [19, 21]. However, approximately 87% of the autopsy samples used in this pleiotropy analysis (as well as in the Beecham et al. study [6]) were from patients with AD. This may indicate that our findings are more relevant with neuropathological progression after onset of AD clinical symptoms. Alternatively, because our study was focused on endophenotypes that might be more proximal than disease diagnosis to effects of the genetic variants [52, 53], our analyses might have more power to detect those novel associations. Finally, to our knowledge, additional large late-onset AD cohorts with neuropathological and genotype data are not currently available for replication of our association findings. Therefore, validation of the role of these loci in AD will likely require experimental evidence.

**Conclusions**

Our findings suggest that genome-wide pleiotropy analysis is a useful approach to identifying novel genetic associations with complex diseases and their endophenotypes. Functional studies are needed to determine whether C2orf40 or HDAC9 is a plausible therapeutic target.

---

**Additional files**

**Additional file 1: Table S1.** Sample characteristics. Table S2. Association P values of Alzheimer disease loci previously established by GWAS in univariate and pleiotropy association tests of neuropathological features. Table S3. Association of cis-eQTL for HDAC9 in the Mayo Clinic brain expression genome-wide association study (eGWS), rs79524815 was not available in the Mayo Clinic brain eGWS, so proxy SNPs that are in LD (r² > 0.90) with rs79524815 were used for the eQTL test with HDAC9 expression. Table S4. Association of expression of SNPs for TRAPPC12-AS1 and AD1 with neuropathological traits and gene expression in the GTEx portal database. Table S5. Association results from the trivariate pleiotropy model of neutitic plaque (NP), neurofibrillary tangles (NFT), and cerebral amyloid angiopathy (CAA) for study-wide significant SNPs in the bivariate model of neuritic plaque (NP), neurofibrillary tangles (NFT), and cerebral amyloid angiopathy (CAA) using the O’Brien method [10]. Figure S1. Quantile-quantile plots of observed (y-axis) vs. expected (x-axis) P values of all SNPs (black dots) and after excluding SNPs in APOE region (blue dots) for the pleiotropy analysis of (a) NP and NFT, (b) NP and CAA, and (c) NFT and CAA using the O’Brien method [10]. Figure S2. Manhattan plots showing genome-wide pleiotropy analyses of (a) NP and NFT, (b) NP and CAA, and (c) NFT and CAA using the O’Brien method [10]. Red dashed line represents the genome-wide significance threshold of P < 5.0 × 10⁻⁸. Loci achieving genome-wide significance are highlighted in red, and known AD genes that attained at least a moderate significance level (P < 10⁻⁷) are highlighted in gold. Figure S3. Regional association plots of genes, including TRAPPC12, TRAPPC12-AS1, and AD1, on chromosome 2 from the joint model of NFT and CAA. Figure S4. Genome-wide trivariate pleiotropy analysis of NP, NFT, and CAA. (a) Quantile-quantile plot. (b) Manhattan plot. (DOCX 1028 kb)

**Acknowledgements**

The authors express appreciation to Dr. Qiiong Yang for help with using the R library (CLUMP package).

**Members of the Alzheimer Disease Genes Consortium**

Marilyn S. Albert1, Roger L. Albin2, Adam Boxer25, James R. Blacker21,22, Bradley F. Boeve23, James D. Bowen24, James R. Carrasquillo32, Steven L. Carroll33, Helena C. Chui34, David G. Clark35, Paul C. Crutch36, Michael M. Crock37, Charles D. Cummings38, F. Yesim Demirci11, Philip De Jager109, Malcolm Dick39, Dennis Dickson36, Jack D. Doupé40, Stanley D. Emory41, Ronald L. Hamilton53, Kara L. Hamilton-Nelson70, John M. Olichney38, Vernon B. sandwiched42, Amanda Myers31, Adam C. Naj14, John M. Olchney38, Vernon S. Pankratz9, Joseph E. Paris87, Amanda Partch54, Henry L. Paulson78, William Perry45, Elaine Peskind51, Ronald C. Petersen34, Alimee Pierce36, Wayne W. Poon77, Huntington Potter89, Joseph F. Quinn34, Ashok Raj55, Murray Raskind32, Eric Reiman15,116, Barry Reisberg44,90, Christiane Reitz23,110,101, John M. Ringman9, Erik D. Robertson45, Ekaterina Rogaeva23.
data analysis. MA, XW, TJM, SGY, DWD, TEG, NDP, and NET provided and JC, GRJ, and LAF wrote the manuscript. JC, XZ, KLL, and GRJ performed the controls. YM, GB, and JM provided technical support. RM, JLH, MAPV, GS, \textit{et al.} Alzheimer's Research & Therapy 2018, 10, 22

Funding
The Alzheimer’s Disease Genetics Consortium supported the collection of samples used in this study through National Institute on Aging (NIA) grants U01-AG032984 and RC2AG036528. Data for this study were prepared, archived, and distributed by the National Institute on Aging Alzheimer’s Disease Data Storage Site (NIAGADS) at the University of Pennsylvania and funded by NIA grant U24-AG041689-01. This work was also supported by NIA grants R01-AG048927 (to BIOP), R01-AG032990 and RF-AG051504 (to NET), and U01-AG046139 (to NET, SGY, TEG, and NDP), as well as by National Institute of Neurological Disorders and Stroke (NINDS) grant R01-NS08820 (to NET). The Brain and Body Donation Program at Sun Health Research Institute (Sun City, AZ, USA) is supported by NINDS grant U24-NS072026 for the National Brain and Tissue Resource for Parkinson’s Disease and Related Disorders; NIA grant P50-AG19610 for the Arizona Alzheimer’s Disease Core Center; Arizona Department of Health Services contract 211002 for the Arizona Alzheimer’s Research Center; Arizona Biomedical Research Commission contracts 4001, 0011, 05-001, and 1001 to the Arizona Parkinson’s Disease Consortium; and the Michael J. Fox Foundation for Parkinson’s Research.

Availability of data and materials
Summarized genome-wide association study results are available from the National Institute on Aging Genetics of Alzheimer’s Disease Data Storage Site (https://www.niagads.org/).

Authors’ contributions
JC, GRI, and LAF wrote the manuscript. JC, XZ, KLL, and GRI performed the data analysis. MA, XW, TJM, SGY, DWD, TEG, NDF, and NET provided and interpreted gene expression data from Alzheimer’s disease cases and controls. YM, GB, and JM provided technical support. RM, JH, MAPV, GS, and LAF obtained the funding for this study. GRI and LAF supervised the project. All authors read and approved the final manuscript.

Ethics approval and consent to participate
This study was approved by the Boston University Institutional Review Board.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Biostatistics, Boston University School of Public Health, Boston, MA, USA. 2Department of Neurology, Mayo Clinic, Jacksonville, FL, USA. 3Department of Health Sciences Research, Mayo Clinic, Jacksonville, FL, USA. 4Department of Pathology, University of Miami Miller School of Medicine, Miami, FL, USA. 5Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA. 6Department of Pathology, University of Washington, Seattle, WA, USA. 7Department of Neurology, Mayo Clinic, Jacksonville, FL, USA. 8Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. 9Department of Neurology, Boston University School of Medicine, Boston, MA, USA. 10Department of Neurology and Geriatrics Center, Columbia University, New York, NY, USA. 11Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA. 12Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA. 13Neurogenetics and Integrated Genomics, Andover Innovative Medicines Institute, Andover, MA, USA. 14Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA. 15Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA.

Received: 27 August 2017 Accepted: 22 January 2018

References
1. Jorm AF, Jolley D. The incidence of dementia: a meta-analysis. Neurology. 1998;51(3):729–33.
2. Conrado MM, Brookmeyer R, Paganini-Hill A, Berlan D, Kawas CH. Dementia incidence continues to increase with age in the oldest old: the 90+ study. Ann Neurol. 2011;67(1):114–21.
3. Shoji M, Golde TE, Ghiso J, Cheung TT, Estus S, Shaffer LM, Cai XD, McKay DM, Tintner R, Frangione B, et al. Production of the Alzheimer amyloid β protein by normal proteolytic processing. Science. 1992;258(5079):125–9.
4. Trojanowski JQ, Schmidt ML, Shin RW, Bramlett GT, Rao D, Lee VM. Altered tau and neurofilament proteins in neuro-degenerative diseases: diagnostic implications for Alzheimer’s disease and Lewy body dementias. Brain Pathol. 1993;3(1):45–54.
5. Love S. Contribution of cerebral amyloid angiopathy to Alzheimer’s disease. J Neurol Neurosurg Psychiatry. 2004;75(1):1–9.
6. Beecham GW, Hamilton K, Naj AC, Martin ER, Huettelmann M, Myers AJ, Conneveaux JJ, Hardy J, Vonsattel JP, Younkin SG, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer’s disease and related dementias. PLoS Genet. 2014;10(9):e1004606.
7. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Burovs J, Gallins PJ, Buixad BM, Jankovic J, Kaukinen PK, et al. Common variants at MS4A4/4A6/8 and CD33 are associated with late-onset Alzheimer’s disease. Nat Genet. 2011;43(13):436–41.
8. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. Acta Neuropathol. 1991;82(4):239–59.
9. O’brien PC. Procedures for comparing samples with multiple endpoints. Biometrics. 1984;40(4):1079–87.
10. Yang Q, Wu H, Guo CY, Fox CS. Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. Genet Epidemiol. 2010;34(5):444–44.
11. Yang Q, Wang Y. Methods for analyzing multivariate phenotypes in genetic association studies. J Probst Stat. 2012;2012:652569.
12. Liu JZ, Mcrae AF, Nyholton DR, Lindel SE, Wray NR, Brown KM, Investigators AMFS, Hayward NK, Montgomery GW, Visscher PM, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010;87(1):39–45.
13. GTEX Consortium. The Genotype-Tissue Expression (GTEX) pilot analysis: multivariate gene regulation in humans. Science. 2015;348(6235):648–60.
14. Zhou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, Carrauqillo MM, Rowley CN, Nair AA, Muddia S, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genet. 2012;8(6):e1002707.
15. Carrauqillo MM, You F, Pankratz VS, Wilcox SL, Ma L, Walker LP, Younkin SG, Younkin CS, Younkin LH, Biscoglio GD, et al. Genetic variation in PCDH7 is associated with susceptibility to late-onset Alzheimer’s disease. Nat Genet. 2009;41(2):192–8.
16. Allen M, Carrauqillo MM, Funk C, Heavner BD, Zou F, Younkin CS, Burgess JD, Chai HS, Crook J, Eddy JA, et al. Human whole genome genotype and transcriptome data for Alzheimer’s disease and other neurodegenerative diseases. Sci Data. 2016;3:160089.
17. Zhang B, Gaiteri C, Bodea LG, Wang Z, McEwen J, Podtelezhnikov AA, Zhang C, Xie T, Tran L, Dobrin R, et al. Integrated systems approach for Alzheimer’s disease genetics.
