Investigation of 15 of the top candidate genes for late-onset Alzheimer’s disease

Olivia Belbin · Minerva M. Carrasquillo · Michael Crump · Oliver J. Culley · Talisha A. Hunter · Li Ma · Gina Bisceglio · Fanggeng Zou · Mariet Allen · Dennis W. Dickson · Neil R. Graff-Radford · Ronald C. Petersen · Kevin Morgan · Steven G. Younkin

Abstract The 12 genome-wide association studies (GWAS) published to-date for late-onset Alzheimer’s disease (LOAD) have identified over 40 candidate LOAD risk modifiers, in addition to apolipoprotein (APOE) e4. A few of these novel LOAD candidate genes, namely BIN1, CLU, CR1, EXOC3L2 and PICALM, have shown consistent replication, and are thus credible LOAD susceptibility genes. To evaluate other promising LOAD candidate genes, we have added data from our large, case–control series (n = 5,043) to meta-analyses of all published follow-up case–control association studies for six LOAD candidate genes that have shown significant association across multiple studies (TNK1, GAB2, LOC651924, GWA_14q32.13, PGBD1 and GALP) and for an additional nine previously suggested candidate genes. Meta-analyses remained significant at three loci after addition of our data: GAB2 (OR = 0.78, p = 0.007), LOC651924 (OR = 0.91, p = 0.01) and TNK1 (OR = 0.92, p = 0.02). Breslow–Day tests revealed significant heterogeneity between studies for GAB2 (p < 0.0001) and GWA_14q32.13 (p = 0.006). We have also provided suggestive evidence that PGBD1 (p = 0.04) and EBF3 (p = 0.03) are associated with age-at-onset of LOAD. Finally, we tested for interactions between these 15 genes, APOE e4 and the five novel LOAD genes BIN1, CLU, CR1, EXOC3L2 and PICALM but none were significant after correction for multiple testing. Overall, this large, independent follow-up study for 15 of the top LOAD candidate genes provides support for GAB2 and LOC651924 (6q24.1) as risk modifiers of LOAD and novel associations between PGBD1 and EBF3 with age-at-onset.

Introduction

Genome-wide association studies represent a powerful approach for identifying putative candidate genes for common complex disorders, such as LOAD. To-date, 12 GWAS of LOAD have been published (Coon et al. 2007; Grupe et al. 2007; Abraham et al. 2008; Bertram et al. 2008; Li et al. 2008; Beecham et al. 2009; Carrasquillo et al. 2009; Harold et al. 2009; Lambert et al. 2009; Poduslo et al. 2009; Potkin et al. 2009; Seshadri et al. 2010) revealing more than 40 candidate variants that modify LOAD risk independent of APOE (and genes likely to be in linkage disequilibrium with APOE such as TOMM40, APOC1 and APOC2). Notably,
CLU is the only signal besides these four APOE-related signals to be identified in more than one GWAS at a genomewide significant level (Harold et al. 2009; Lambert et al. 2009).

Follow-up case–control association studies of the 40+ candidate loci identified by GWAS are vital in order to filter out any false-positive signals and to provide further evidence for genetic association of the truly functional genes. We have successfully replicated (Carrasquillo et al. 2009; Lambert et al. 2009) the association of variants in CLU, CR1 and PI-CALM identified by two large GWAS (Harold et al. 2009; Lambert et al. 2009) thus providing compelling support for these genes as true candidate LOAD genes. Two other recently GWAS-identified signals in EXOC3L2 and BIN1 (Seshadri et al. 2010) are currently being investigated for replication in our large case–control series.

In addition to these candidates, AlzGene (http://www.Alzgene.org) meta-analyses of all published LOAD association studies have revealed significant association for six other GWAS-identified variants (TNK1, GAB2, LOC651924, GWA_14q32.13, PGBD1 and GALP) that are now ranked among the Top 50 LOAD candidate genes on the AlzGene website (Bertram et al. 2006). Although, many variants have yet to be tested in follow-up studies, these six loci currently represent compelling GWAS signals worthy of follow-up investigation (Bertram and Tanzi 2009).

Here we evaluate in our large case–control series (n = 5,043) the most significant variants in TNK1, GAB2, LOC651924, GWA_14q32.13, PGBD1 and GALP for genetic association with LOAD, as well as nine previously suggested candidate LOAD genes (Grupe et al. 2007). We have performed meta-analyses of all available published case–control series including our data and investigated the effect of heterogeneity on the ORs between series. We also tested for association of these 15 variants with age-at-onset of LOAD and for epistatic interaction with APOE ε4, BIN1, CLU, CR1, EXOC3L2 and PICALM.

Methods

Case–control subjects

The case–control series consisted of 5,043 Caucasian subjects from the United States (2,455 AD, 2,588 control) ascertained at the Mayo Clinic (1,753 AD, 2,379 controls) or through the National Cell Repository for Alzheimer’s Disease (NCRAD: 702 AD, 209 control). All subjects ascertained at the Mayo Clinic in Jacksonville, Florida (JS: 602 AD, 604 control) and at the Mayo Clinic in Rochester, Minnesota, (RS: 553 AD, 1,399 control) were diagnosed by a Mayo Clinic neurologist. The neurologist confirmed a clinical dementia rating score of 0 for all JS and RS subjects enrolled as controls; cases had diagnoses of possible or probable AD made according to NINCDS-ADRDA criteria (McKhann et al. 1984). In the autopsy-confirmed series (AUT: 598 AD, 376 control), all brains were evaluated by Dr. Dennis Dickson and came from the brain bank maintained at the Mayo Clinic in Jacksonville, FL. In the AUT series the diagnosis of definite AD was also made according to NINCDS-ADRDA criteria. All AD brains analyzed in the study had a Braak score of 4.0 or greater. Brains employed as controls had a Braak score of 2.5 or lower but often had brain pathology unrelated to AD and pathological diagnoses that included vascular dementia, fronto-temporal dementia, dementia with Lewy bodies, multi-system atrophy, amyotrophic lateral sclerosis, and progressive supranuclear palsy.

One AD case from each of the 702 late-onset NCRAD families was analyzed. NCRAD AD cases were selected based on strength of diagnosis (autopsy-confirmed, 32% > probable, 45% > possible, 8% > family report, 15%); the case with the earliest age at diagnosis was taken when several cases had equally strong diagnoses. The 209 NCRAD controls that we employed were unrelated Caucasian subjects from the United States with a clinical dementia rating of 0, specifically collected for inclusion in case–control series. All individuals with an age-at-diagnosis <60 or with mutations in PSEN1, PSEN2 or APP were removed from analyses. The mean age-at-diagnosis, percentage that are female and percentage that possess at least one copy of the APOE ε4 allele for each series are shown in Online Resource 1.

DNA isolation

For the JS and RS samples, DNA was isolated from whole blood using an AutoGen instrument (AutoGen, Inc, Holliston, MA). The DNA from AUT samples was extracted from cerebellum using Wizard® Genomic DNA Purification Kits (Promega Corp., Madison, WI). DNA from the RS and AUT series was scarce, so samples from these two series were subjected to whole genome amplification using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences Corp., Piscataway, NJ).

Genotyping of variants

All genotyping was performed on 384-well plate formats containing on average eight (min = 4, max = 14) negative controls per plate. We ensured that each plate had a mixture of cases and controls. Positive controls were not included on these plates. GAB2 and PGBD1 variants were genotyped at the Mayo Clinic in Jacksonville using TaqMan® SNP Genotyping Assays in an ABI PRISM® 7900HT Sequence Detection System with 384-Well Block Module from Applied Biosystems, California, USA. The genotype data was analyzed using the SDS software version 2.2.2.
Genotype information for BIN1 (rs744373), EXOCS3L2 (rs597668), MYH13 (rs2074877), PCK1 (rs8192708) and TRAK2 (rs1302344), were available from our GWAS (Carrasquillo et al. 2009). All other variants were genotyped using SEQUENOM’s MassArray iPLEX technology (SEQUENOM Inc, San Diego, CA) following the manufacturer’s instructions. Genotype calls were made using the default post-processing calling parameters in SEQUENOM’s Typer 4.0 software, followed by visual inspection to remove genotype calls that were obviously erroneous based on the presence or absence of allele peaks in an individual sample’s spectrogram, to check that the boundaries of the genotype clusters were non-overlapping and finally to ensure that samples between clusters were not called. Genotyping probe sequences are non-overlapping and finally to ensure that samples between clusters were not called. Genotyping probe sequences are shown in Online Resource 2. All variants passed the p value cut-off (p > 0.001) for deviation from Hardy–Weinberg equilibrium as suggested by Wigginton et al. when investigating >1,000 samples (Wigginton et al. 2005). Genotyping of variants in CLU, PICALM and CR1 have been reported previously (Carrasquillo et al. 2010).

Statistical analyses

All statistics were performed using StatsDirect v2.5.8 software. Variants were analyzed for association with LOAD by logistic regression (additive/allelic dosage, dominant and recessive models). When included in the analysis, covariates were sex, age at diagnosis/entry, and APOE ε4+/−. Meta-analyses were performed for each individual series including all available published case–control series from Caucasian populations for these variants. Genotype counts for previously published studies were obtained either directly from the publication or from the AlzGene website. Summary ORs and 95% CI were calculated using the DerSimonian and Laird (1986) random-effects model. Breslow–Day tests were used to test for heterogeneity between series. Since age-at-onset of LOAD did not follow a Gaussian distribution, association of the minor allele at each variant with age-at-onset was performed using a Mann–Whitney U test. Tests for epistatic interaction were performed using the Synergy Factor Excel spreadsheet made available by Cortina-Borja et al. (2009). Odds ratios were calculated based on a dominant model for the minor allele, i.e. major allele homozygotes versus heterozygotes and minor allele homozygotes and, in the case of APOE ε4, no-ε4 versus at least one copy of ε4.

Results

We have genotyped 15 candidate LOAD variants initially identified by GWA studies. Genotype and allele counts for these variants in our complete series are shown in Online Resource 3. We first focused on six variants highlighted by Bertram and Tanzi (2009) as the most compelling GWAS signals worthy of follow-up investigation. The initial GWAS findings for these six variants are shown in Table 1a. In order to directly compare our results with these previous studies, we tested for association with LOAD in our case–control series by logistic regression using an additive/allelic dosage model (Table 1b). Although all variants showed significant association with LOAD (all p < 0.0003) in the initial studies, no variants were significantly associated with LOAD risk in our study (all p > 0.07). Furthermore, only two variants had ORs in the same direction [GAB2; Reiman OR = 0.55 (Reiman et al. 2007), Mayo OR = 0.94 and LOC651924; Grupe OR = 0.86 (Grupe et al. 2007), Mayo OR = 0.94]. It must be noted that the Mayo series were larger in number than the initial studies and therefore, in principle, had greater power to detect these associations.

In order to determine whether the data from our series would confound the significant meta-analysis reported by AlzGene, we performed our own meta-analyses including our data using the DerSimonian and Laird random-effects model (Table 2). Given that none of the six variants highlighted in the Bertram and Tanzi (2009) showed significant association with risk of LOAD in our case–control series, it is not surprising that the overall ORs for all six variants were closer to 1 when our data (Table 2b) were included (Table 2c) compared to the meta-analyses of all previous studies (Table 2a). However, while significance was diminished following inclusion of our data, the meta-analyses for TNK1 (p = 0.02), GAB2 (p = 0.007) and LOC651924 (p = 0.01) remained significant at the p < 0.05 level. The remaining three loci, though not significant had overall ORs in the same direction as the AlzGene meta-analysis when our data were included.

In order to investigate heterogeneity between studies for these variants, Breslow–Day tests were performed (shown in Table 2; Fig. 1). Forest plots of the OR and 95% CI for each series are shown in Fig. 1. Two variants (GAB2, p < 0.0001; GALP, p = 0.03) showed significant heterogeneity in the previously published data (Table 2a), while in our data (Table 2b) only GAB2 showed significant heterogeneity (p = 0.0002). Overall (Table 2c; Fig. 1), GAB2 showed the most significant heterogeneity (p < 0.0001) followed by GWA_14q32.13 (p = 0.006) and GALP (p = 0.05).

Since some genetic variants may exert dominant or recessive effects we also performed logistic regression using these models and corrected for sex, age at diagnosis/entry, and APOE ε4+/− as covariates (Online Resource 4). Although LOC651924 (OR = 0.85, p = 0.03) and GWA_14q32.13 (OR = 1.18, p = 0.01) gave significant ORs under a dominant model, neither would survive Bonferroni correction for the 60 tests performed.
Table 1 Replication results for genetic association of six candidate LOAD loci identified by GWAS

| Gene       | Variant | Study (# pops) | Total N (AD:CON) | MAF (AD:CON) | OR   | 95% CI     | p          |
|------------|---------|----------------|------------------|--------------|------|------------|------------|
| (a) Initial study |
| **TNK1**  | rs1554948 | Grupe (5)   | 3,913 (1,828:2,085) | 0.454:0.497 | 0.84\(^a\) | 0.77–0.92 | 6 × 10\(^{-5}\) |
| **GAB2**  | rs10793294 | Reiman (3)  | 1,411 (644:767)  | 0.190:0.300 | 0.55\(^a\) | 0.46–0.65 | 1 × 10\(^{-7}\) |
| **LOC65192 4** | rs6907175 | Grupe (5)   | 3,913 (1,828:2,085) | 0.461:0.498 | 0.86\(^a\) | 0.77–0.96 | 3 × 10\(^{-4}\) |
| **GWA_14q32.13** | rs11622883 | Grupe (5)   | 3,913 (1,828:2,085) | 0.422:0.465 | 0.84\(^a\) | 0.77–0.93 | 9 × 10\(^{-5}\) |
| **PGBD1** | rs3800324  | Grupe (5)   | 3,913 (1,828:2,085) | 0.048:0.034 | 1.43 | 1.13–1.80 | 3 × 10\(^{-4}\) |
| **GALP**  | rs3745833  | Grupe (5)   | 3,913 (1,828:2,085) | 0.387:0.345 | 1.20 | 1.09–1.32 | 5 × 10\(^{-5}\) |
| (b) Mayo   |
| **TNK1**  | rs1554948 | Mayo (4)     | 5,043 (2,455:2,588) | 0.453:0.448 | 1.02 | 0.94–1.10 | 0.62 |
| **GAB2**  | rs10793294 | Mayo (4)     | 5,043 (2,455:2,588) | 0.214:0.224 | 0.94 | 0.86–1.03 | 0.20 |
| **LOC65192 4** | rs6907175 | Mayo (4)     | 5,043 (2,455:2,588) | 0.487:0.473 | 0.94 | 0.87–1.02 | 0.15 |
| **GWA_14q32.13** | rs11622883 | Mayo (4)     | 5,043 (2,455:2,588) | 0.455:0.437 | 1.07 | 0.99–1.16 | 0.07 |
| **PGBD1** | rs3800324  | Mayo (4)     | 5,043 (2,455:2,588) | 0.043:0.047 | 0.91 | 0.75–1.10 | 0.32 |
| **GALP**  | rs3745833  | Mayo (4)     | 5,043 (2,455:2,588) | 0.365:0.364 | 1.00 | 0.93–1.09 | 0.93 |

Association results from logistic regression using an additive/allelic dosage model and no correction for covariates for each variant are shown from (a) the initially published study and from (b) our data. Study; first author named on initial publication; (Grupe et al. 2007; Reiman et al. 2007)

\# pops number of independent series tested, MAF minor allele frequency

\(^a\) OR is shown for minor allele whereas the initial study reported association for the major allele

\(^b\) Excludes monomorphic series UK3

Table 2 Meta-analysis of the six loci for association with LOAD in

(a) previously published studies (Grupe et al. 2007; Reiman et al. 2007; Li et al. 2008; Feulner et al. 2009; Figgins et al. 2009; Sleegers et al. 2009) (b) our data and (c) overall; including all published studies in Caucasian series and our data

| Gene       | Variant | # pops. | Total N (AD:CON) | OR   | 95% CI | p          | Breslow–Day p |
|------------|---------|---------|------------------|------|--------|------------|---------------|
| (a) AlzGene meta-analysis |
| **TNK1**  | rs1554948 | 6       | 5,932 (2,837:3,095) | 0.86 | 0.80–0.93 | 0.0002       | 0.37          |
| **GAB2**  | rs10793294 | 5       | 4,029 (2,142:1,887) | 0.69 | 0.54–0.88 | 0.003        | 0.0009       |
| **LOC65192 4** | rs6907175 | 6       | 5,932 (2,837:3,095) | 0.89 | 0.82–0.96 | 0.005        | 0.39          |
| **GWA_14q32.13** | rs11622883 | 6       | 5,932 (2,837:3,095) | 0.88 | 0.80–0.97 | 0.01         | 0.18          |
| **PGBD1** | rs3800324  | 7       | 10,815 (5,156:5,659) | 1.21 | 1.02–1.44 | 0.03         | 0.53          |
| **GALP**  | rs3745833  | 6       | 5,932 (2,837:3,095) | 1.13 | 1.00–1.29 | 0.06         | 0.03          |
| (b) Mayo data meta-analysis |
| **TNK1**  | rs1554948 | 4       | 5,043 (2,455:2,588) | 1.00 | 0.92–1.09 | 0.99         | 0.61          |
| **GAB2**  | rs10793294 | 4       | 5,043 (2,455:2,588) | 0.89 | 0.68–1.16 | 0.40         | 0.0002       |
| **LOC65192 4** | rs6907175 | 4       | 5,043 (2,455:2,588) | 0.94 | 0.83–1.07 | 0.37         | 0.08          |
| **GWA_14q32.13** | rs11622883 | 4       | 5,043 (2,455:2,588) | 1.09 | 0.99–1.20 | 0.07         | 0.07          |
| **PGBD1** | rs3800324  | 4       | 5,043 (2,455:2,588) | 0.87 | 0.70–1.06 | 0.17         | 0.76          |
| **GALP**  | rs3745833  | 4       | 5,043 (2,455:2,588) | 1.02 | 0.94–1.12 | 0.64         | 0.39          |
| (c) Overall meta-analysis |
| **TNK1**  | rs1554948 | 10      | 10,975 (5,272:5,932) | 0.92 | 0.85–0.98 | 0.02         | 0.13          |
| **GAB2**  | rs10793294 | 9       | 9,072 (4,597:4,475)  | 0.78 | 0.64–0.93 | 0.007        | <0.0001      |
| **LOC65192 4** | rs6907175 | 10      | 10,975 (5,272:5,932) | 0.91 | 0.85–0.98 | 0.01         | 0.15          |
| **GWA_14q32.13** | rs11622883 | 10      | 10,975 (5,272:5,932) | 0.96 | 0.88–1.06 | 0.44         | 0.006         |
| **PGBD1** | rs3800324  | 11      | 15,858 (7,611:8,247) | 1.07 | 0.92–1.24 | 0.36         | 0.25          |
| **GALP**  | rs3745833  | 10      | 10,975 (5,272:5,932) | 1.08 | 0.99–1.17 | 0.07         | 0.05          |

# pops number of independent series tested, Breslow–Day p p value for series heterogeneity Breslow–Day tests
We next tested for association of the variants with age-at-onset in the 2,455 LOAD patients from our case–control series (Table 3). The only variant to show association at the $p < 0.05$ level was $PGBD1$ ($p = 0.04$) where the minor allele was associated with an age-at-onset 1 year earlier than the major allele, although this weak association would not survive Bonferroni correction ($p < 0.003$).

In addition to these six loci, we had genotype information available for nine variants ($EBF3$, $LMNA$, $BCR$, $UBD$, $THEM5$, $CTSS$, $TRAK2$, $MYH13$ and $PCK1$) identified by Grupe et al. (2007). The genotype counts, case–control association, meta-analyses, alternative models, association with age-at-onset and epistatic interactions can be found in Online Resources 3–8. In summary, although all nine variants were associated with LOAD in the initial study (all $p < 0.001$), logistic regression of our data using an additive/allelic dosage model and no covariates revealed one variant (rs13022344 in $TRAK2$) significantly associated with LOAD in our series ($OR = 0.86$, $p = 0.02$) but in the opposite direction to that reported in the initial study ($OR = 1.07$, $p = 0.001$). Notably, $EBF3$ ($p = 0.04$), $THEM5$ ($p = 0.03$), $CTSS$ ($p = 0.03$) and $TRAK2$ ($p = 0.02$) were associated with LOAD risk under a recessive model whilst correcting for covariates, although these associations would not survive Bonferroni correction for the 60 tests performed ($p < 0.0008$). None of these nine

| Gene   | Variant     | N  | Mean (SD) | U     | p value |
|--------|-------------|----|-----------|-------|---------|
|        |             |    | Maj       | Min   |         |
| $TNK1$ | rs1554948   | 2,442 | 78.3 (7.7) | 78.4 (7.7) | 616,938.5 | 0.85 |
| $GAB2$ | rs10793294  | 2,416 | 78.4 (7.6) | 78.5 (7.7) | 692,338.5 | 0.68 |
| $LOC651924$ | rs6907175 | 2,443 | 78.8 (7.9) | 78.3 (7.6) | 520,939.0 | 0.12 |
| $GWA_{14q32.13}$ | rs11622883 | 2,447 | 78.4 (7.3) | 78.4 (7.8) | 623,344.0 | 0.93 |
| $PGBD1$ | rs3800324 | 2,435 | 78.5 (7.6) | 77.5 (7.7) | 244,760.0 | 0.04 |
| $GALP$ | rs3745833  | 2,446 | 78.4 (7.7) | 78.0 (7.7) | 339,685.0 | 0.43 |

Degrees of freedom ($df$) and number of samples included in the analyses ($N$) are given for each variant. Mean age-at-onset and standard deviation (SD) are shown for each group defined by possession of two copies of the major allele (Maj) or at least one copy of the minor allele (Min). The $p$ value provides the significance level for the $U$ statistic from a Mann–Whitney $U$ test comparing mean age-at-onset between the two groups.
variants were significant at the \( p < 0.05 \) level following meta-analyses either before or after addition of our data (Online Resources 6). Overall, Breslow–Day tests revealed genetic heterogeneity for six (\( LMNA \ p = 0.0004, BCR \ p = 0.04, THEM5 \ p = 0.01, PCK1 \ p = 0.003, CTSS \ p = 0.02 \) and \( TRAK2 \ p = 0.0002 \)) of the nine variants. None of these nine variants were associated with age-at-onset (Online Resources 7) after Bonferroni correction, however, \( EBF3 \) showed nominally significant association with a later age-at-onset of 0.8 years (79.0 years) compared to the major allele (78.2 years; \( p = 0.03 \)).

In order to determine whether these 15 variants interact with other strong LOAD candidates to modify risk for LOAD, we tested for epistatic interaction between the variants studied here and the strongest known LOAD risk factor, \( APOE \) ε4, as well as the top GWAS-identified variants for which we had genotype information available: \( BIN1 \) (rs744373), \( CLU \) (rs11136000), \( CRI \) (rs3818361), \( EXOC3L2 \) (rs597668) and \( PICALM \) (rs3851179). The results for all 105 tests performed are shown in Online Resource 8. There were seven interactions that were significant at the \( p \leq 0.05 \) level, which are shown in Table 4, however, none would survive Bonferroni correction for the 105 tests performed. Further investigation of these possible epistatic interactions in multiple, independent studies is required in order to determine whether there is true synergy between the variants.

**Discussion**

This study used the largest sample size to-date to investigate 15 of the top AlzGene hits (AlzGene, accessed October 2010) which were originally identified in a LOAD GWAS. Meta-analyses remained significant at three loci after addition of our data: \( GAB2 \) (rs10793294, \( OR = 0.78, p = 0.007 \)), \( LOCC51924 \) (rs6907175, \( OR = 0.91, p = 0.01 \)) and \( TNK1 \) (rs1554948, \( OR = 0.92, p = 0.02 \)). Although our data alone provided no support for an association of \( TNK1 \) with LOAD (\( OR = 1.00, p = 0.99 \)), the AlzGene meta-analyses odds ratios for both \( GAB2 \) (0.69) and \( LOCC51924 \) (0.89) were well replicated in our series (\( OR = 0.89 \) and 0.94, respectively) albeit that neither variant was significant (\( p = 0.40 \) and 0.37, respectively). We also investigated nine additional variants (\( EBF3, LMNA, BCR, THEM5, PCK1, MYH13, CTSS, UBD \) and \( TRAK2 \)) identified by Grupe et al. but found no significant associations following meta-analyses.

Our meta-analyses which include nine independent studies comprising 9,072 individuals provide good evidence that \( GAB2 \) (rs10793294 \( OR = 0.78, p = 0.007 \)) is a genuine candidate LOAD locus. These data are further supported by a recent family-based study (Schjeide et al. 2009a), which revealed significant association of another \( GAB2 \) variant (rs7101429) in 399 families (\( p = 0.002 \)), thus strengthening the evidence for \( GAB2 \). In consideration of this association of \( GAB2 \) with LOAD in families, we performed our logistic regression (additive model) analyses again on the total dataset with the 112 NCRAD LOAD patients with a family history of AD removed. We found that removing these samples gave a comparable association to our initial analyses (all samples: \( n = 4,969, OR = 0.94, p = 0.20 \); no family history: \( n = 4,857, OR = 0.95, p = 0.36 \); data not shown). We also found a comparable association of \( GAB2 \) with age-at-onset removal of these samples (all samples: \( n = 2,416, U = 692,338.5, p = 0.68 \); no family history: \( n = 2,304, U = 628,587, p = 0.76 \); data not shown).

Notably, in another family-based study by the same group, eight variants included in this manuscript (\( GALP, GWA_{14q32.13}, LMNA, LOCC51924, MYH13, PCK1, PGBD1, TNK1 \)) were tested but failed to show association with LOAD in 457 families (Schjeide et al. 2009b); this is compatible with our meta-analyses of variants in \( GWA_{14q32.13} \) (rs11622883), \( PGBD1 \) (rs3800324) and \( GALP \) (rs3745833) which revealed no association with LOAD.

Our meta-analyses also provided evidence that \( LOCC51924 \) is a true candidate locus (\( OR = 0.91, p = 0.01 \)). Although only one out of the nine series studied revealed significant association (\( p < 0.05 \)), the effect of the variants were in the same direction (with comparable ORs) in seven of the series. As a result, the meta-analysis revealed significant association thus supporting the evidence for \( LOCC51924 \) as a LOAD candidate.

The 15 variants we analyzed showed remarkable across-study heterogeneity. The Breslow–Day \( p \) values for the initial, Mayo follow-up and overall meta-analyses of the 15 variants we analyzed are summarized in Table 2 and Online Resource 6. Overall, meta-analysis of the variants in four genes (\( GAB2, TRAK2, LMNA \) and \( PCK1 \)) gave Breslow–Day \( p \) values ranging from <0.0001 to 0.002 that are significant even after Bonferroni correction for 15 variants analyzed (\( p < 0.003 \)). The variants in seven genes had nominally significant or highly suggestive Breslow–Day \( p \) values that ranged from 0.01 to 0.06, and the variants in the three remaining genes had Breslow–Day \( p \) values of 0.12 to 0.25. Thus, our analysis of 15 promising LOAD variants suggests that LOAD variants may often show noteworthy series to series heterogeneity. If the heterogeneity we observed is real and if it occurs as frequently as our data suggest, then many genetic variants may influence LOAD susceptibility in a way that depends on genetic and/or environmental factors that vary from series to series.

It is now clear that, apart from the well-known \( APOE \) alleles, common genetic variants have only weak association with LOAD. Whether many of these variants have
| v1  | v2  | (a) TRAK2 × CR1 | (b) TRAK2 × BIN1 | (c) LOC651924 × PICALM | (d) EXOC3L2 × APOE |
|-----|-----|-----------------|-----------------|------------------------|-------------------|
|     |     | LOAD CON OR     | LOAD CON OR     | LOAD CON OR           | LOAD CON OR       |
| −   | −   | 265 344 Ref     | 235 474 Ref     | 174 79 Ref            | 724 1,616 Ref     |
| +   | −   | 254 459 0.72    | 137 182 1.52    | 506 356 0.65          | 123 311 0.88      |
| −   | +   | 134 205 0.85    | 283 341 1.67    | 192 175 0.50          | 1,324 531 5.57    |
| +   | +   | 185 239 1.00    | 160 220 1.47    | 635 553 0.52          | 253 74 7.63       |
| Total |     | 838 1,247       | 815 1,217       | 1,507 1,163           | 2,424 2,532       |
| Synergy factor |     | 1.65 (1.1–2.4); p = 0.007 | 0.58 (0.4–0.8); p = 0.004 | 1.62 (1.1–2.4); p = 0.01 | 1.55 (1.1–2.2); p = 0.02 |

| v1  | v2  | (e) EXOC3L2 × PICALM | (f) UBD × EXOC3L2 | (g) TRAK2 × EXOC3L2 |
|-----|-----|----------------------|-------------------|-------------------|
|     |     | LOAD CON OR          | LOAD CON OR       | LOAD CON OR       |
| −   | −   | 257 323 Ref          | 142 270 Ref       | 249 389 Ref       |
| +   | −   | 138 180 0.96         | 382 559 1.30      | 281 444 0.99      |
| −   | +   | 275 512 0.68         | 99 120 1.57       | 147 157 1.46      |
| +   | +   | 170 226 0.95         | 204 284 1.37      | 158 247 1.00      |
| Total |     | 840 1,241           | 827 1,233         | 835 1,237        |
| Synergy factor |     | 1.45 (1.0–2.1); p = 0.05 | 0.67 (0.4–1.0); p = 0.05 | 0.69 (0.5–1.0); p = 0.05 |

Genotype counts are given for LOAD patients and controls (CON). Counts are stratified by each combination of two variants; for each combination, variant 1 (v1) is the variant in the gene listed first and (v2) is the variant in the gene listed second; The total counts for LOAD patients and controls are given underneath the stratified counts. Odds ratios (OR) for each stratified group compared to the referent OR, synergy factor with 95% confidence intervals in parentheses and p value for the interaction are also provided; Variants are as follows; TRAK2:rs1302234, CR1:rs3818361, BIN1:rs744373, LOC651924:rs6907175, PICALM:rs3851179, EXOC3L2:rs597668, APOE:e4, PICALM:rs3851179, UBD:rs444013

Ref OR for individuals carrying the major allele at both variants

+ refers to individuals carrying at least one copy of the minor allele at that variant, – refers to individuals carrying the major allele only
odds ratios that truly vary because they depend on environmental and/or genetic factors that differ from series to series is currently unclear. What is clear is that variants of this type are likely to be missed if genetic association studies focus exclusively on replicable associations that become highly significant when many series are combined. To find important susceptibility alleles with effects that vary from series, it may be necessary to consider and to understand variants that show significant association in some series and highly significant heterogeneity on meta-analysis even though meta-analysis provides no evidence for association.

It is important to recognize that spurious heterogeneity can occur owing to publication bias wherein only those series that, by chance, have false-positive results are published. When these series are combined with follow-up series with ORs that vary randomly around 1.0, Breslow–Day testing can show significant, but misleading evidence of heterogeneity. One way to mitigate this problem is to eliminate all series that have been replicated. When these series are combined with follow-up studies show heterogeneity. In the current study, the 15 variants evaluated in this study as well as with APOE ε4, BIN1 (rs744373), CLU (rs11136000), CR1 (rs3818361), EXOC3L2 (rs597668) and PICALM (rs3851179). Seven pairs showed nominally significant synergy factors (p values ranging from 0.007 to 0.05), but none remained significant after correction for the 105 tests performed (p < 0.0005). It is possible that many epistatic interactions exist between LOAD genes, of which relatively few combinations have been tested here. We therefore propose that future studies of candidate LOAD genes apply tests for epistasis with other candidate genes in order to identify otherwise hidden interactions that could contribute greater risk than any gene individually.

Overall, this study represents a thorough, independent follow-up study of 15 of the top LOAD candidate genes, in a large case–control series and provides further evidence for the association of GAB2 and LOC651924 (6q24.1) with LOAD. In addition, we have provided suggestive evidence that, in our series, two genes (PGBD1 and EBF3) are potentially associated with age-at-onset of LOAD.

The experiments described in this manuscript comply with the current laws of the United States of America where they were performed. Approval was obtained from the ethics committee or institutional review board of each institution responsible for the ascertainment and collection of samples (Mayo Clinic College of Medicine, Jacksonville, FL and Mayo Clinic College of Medicine, Rochester, MN, USA). Written informed consent was obtained for all individuals that participated in this study.

Acknowledgments Samples from the National Cell Repository for Alzheimer’s Disease (NCRAD) were used in this study. We thank contributors, including the Alzheimer’s Disease Centers who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. This work was supported by grants from the US National Institutes of Health, NIA R01 AG18023 (N.R.G.-R., Steven G. Younkin); Mayo Alzheimer’s Disease Research Center, P50 AG16574 (R.C.P., D.W.D., N.R.G.-R., Steven G. Younkin); Mayo Alzheimer’s Disease Patient Registry, U01 AG06576 (R.C.P.); and US National Institute on Aging, AG25711, AG17216, AG03949 (D.W.D.). Samples from the National Cell Repository for Alzheimer’s Disease (NCRAD), which receives government support under a cooperative agreement grant (U24AG21886) awarded by the National Institute on Aging (NIA), were used in this study. This project was also generously supported by the Robert and Clarice Smith Postdoctoral Fellowship (M.M.C.); Robert and Clarice Smith and Abigail Van Buren Alzheimer’s Disease Research Program (R.C.P., D.W.D., N.R.G.-R.; Steven G. Younkin) and by the Palumbo Professorship in Alzheimer’s Disease Research (Steven G. Younkin). K.M. is funded by the Alzheimer’s Research Trust and the Big Lottery Fund.

Conflict of interest The authors declare that they have no conflict of interest.
Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O’Donovan MC, Williams J, Owen MJ, Kirov G (2008) A genome-wide association study for late-onset Alzheimer’s disease using DNA pooling. BMC Med Genomics 1:44

Beecham GW, Martin ER, Li YJ, Slifer MA, Gilbert JR, Haines JL, Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Gerrish A, Hamshere ML, Pahlwa JS, Moskvina V, Dowzell K, Williams K, Jones N, Thomas C, Streiton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinshtein DC, Gill M, Lawlor B, Lovestone S, Holmans P, O’Donovan M, Morris JC, Thal L, Goate A, Owen MJ, Williams J (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(8):865–873

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahlwa JS, Moskvina V, Dowzell K, Williams K, Jones N, Thomas C, Streiton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinshtein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schurrman B, van den Bussche H, Heuser L, Kornhuber J, Wiltfang I, Dichgans M, Frolich L, Hampel H, Hull M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleekers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NG, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsalaki M, Singleton AB, Guererro R, Muhleisen TW, Nothen MM, Moebus S, Jocket KL, Klopp N, Wichmann HE, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O’Donovan M, Owen MJ, Williams J (2009) Genome-wide association study identifies variants atCLU and PICALM associated with Alzheimer’s disease. Nat Genet 41(10):1088–1093

Lambert JC, Heath S, Even G, Campion D, Sleekers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Leteenneur L, Bettens K, Berr C, Passquier F, Fievet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helsalimi S, Porcellini E, Hanon O, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, Manuco M, Panza F, Naemias B, Bossu P, Piccardi P, Annoni G, Seripa D, Galimberti D, Hannequin D, Licastro F, Soininen H, Ritchie K, Blanche H, Dartigues JF, Tzourio C, Gut I, Van Broeckhoven C, Alperovitch A, Lathrop M, Amouyel P (2009) Genome-wide association study identifies variants atCLU and CR1 associated with Alzheimer’s disease. Nat Genet 41(10):1094–1099

Li H, Wettten L, Li L, St Jean PL, Upmanyu R, Surl H, Hosford D, Barnes MR, Briley JD, Borrie M, Coletta N, Delisle R, Dhallia D, Ehnh MG, Feldman HH, Fornazzari L, Gauthier S, Goodgame N, Guzman D, Hammond S, Hollingworth P, Hsiung GY, Johnson J, Kelly DD, Keren R, Kertesz A, King KS, Lovestone S, Lov-English I, Matthews PM, Owen MJ, Plumpton M, Pryse-Phillips W, Prinjha RK, Richardson JC, Saunders A, Slater AI, St George-Hyslop PH, Snittsw SW, Swartz JE, Taylor RL, Wherrret J, Williams J, Yarnall DP, Gibson RA, Irizarry MC, Middleton LT, Roses AD (2008) Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol 65(1):45–53

McKhan G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (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(7):939–944

Poduslo SE, Huang R, Huang J, Smith S (2009) Genome screen of late-onset Alzheimer’s disease extended pedigrees identifies TRPC4AP by haplotype analysis. Am J Med Genet B Neuropsychiatr Genet 150B(1):50–55

Potkin SG, Guftanti G, Lokatos A, Turner J, Krugell F, Fallon JH, Saykin AJ, Oro A, Lupoli S, Salvi E, Weiner M, Macciardi F (2009) Hippocampal atrophy as a quantitative trait in a genomewide association study identifying novel susceptibility genes for Alzheimer’s disease. PLoS One 4(8):e6501

Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD,Pearson JV, Hu-Lince D, Huentelmen MJ,
Craig DW, Coon KD, Liang WS, Herbert RH, Beach T, Rohrer KC, Zhao AS, Leung D, Bryden L, Marlowe L, Kaleem M, Mastroeni D, Grover A, Heward CB, Ravid R, Rogers J, Hutton ML, Melquist S, Petersen RC, Alexander GE, Caselli RJ, Kukull W, Papassotriopoulos A, Stephan DA (2007) GAB2 alleles modify Alzheimer’s risk in APOE epsilon4 carriers. Neuron 54(5):713–720
Schjeide BM, Hooli B, Parkinson M, Hogan MF, DiVito J, Mullin K, Blacker D, Tanzi RE, Bertram L (2009a) GAB2 as an Alzheimer disease susceptibility gene: follow-up of genomewide association results. Arch Neurol 66(2):250–254
Schjeide BM, McQueen MB, Mullin K, DiVito J, Hogan MF, Parkinson M, Hooli B, Lange C, Blacker D, Tanzi RE, Bertram L (2009b) Assessment of Alzheimer’s disease case–control associations using family-based methods. Neurogenetics 10(1):19–25
Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, Bis JC, Smith AV, Carrasquillo MM, Lambert JC, Harold D, Schrijvers EM, Ramirez-Lorca R, Debette S, Longstreth WT Jr, Janssens AC, Pankratz VS, Dartigues JF, Hollingworth P, Aspelund T, Hernandez I, Beiser A, Kuller LH, Koudstaal PJ, Dickson DW, Tzourio C, Abraham R, Antunez C, Du Y, Rotter JJ, Aulchenko YS, Harris TB, Petersen RC, Berr C, Owen MJ, Lopez-Arrieta J, Varadarajan BN, Becker JT, Rivadeneira F, Nalls MA, Graff-Radford NR, Campion D, Auerbach S, Rice K, Hofman A, Jonsson PV, Schmidt H, Lathrop M, Mosley TH, Au R, Psaty BM, Uitterlinden AG, Farrer LA, Lumley T, Ruiz A, Williams J, Amouyel P, Younkin SG, Wolf PA, Launer LJ, Lopez OL, van Duijn CM, Breteler MM (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 303(18):1832–1840
Sleegers K, Bettens K, Brouwers N, Engelborghs S, van Mieghem H, De Deyn PP, Van Broeckhoven C (2009) Common variation in GRB-associated binding protein 2 (GAB2) and increased risk for Alzheimer disease. Hum Mutat 30(2):E338–E344
Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy–Weinberg equilibrium. Am J Hum Genet 76:887–893