Meta-analysis of genetic association with diagnosed Alzheimer's disease identifies novel risk loci and implicates Abeta, Tau, immunity and lipid processing

Introduction (word count: 149 of 150 word limit)

Late-onset Alzheimer's disease (LOAD, onset age > 60 years) is the most prevalent dementia in the elderly[1], and risk is partially driven by genetics[2]. Many of the loci responsible for this genetic risk were identified by genome-wide association studies (GWAS)[3–8]. To identify additional LOAD risk loci, we performed the largest GWAS to date (92,255 individuals), analyzing both common and rare variants. We confirm 20 previous LOAD risk loci and identify four new genome-wide loci (*IQCK*, *ACE*, *ADAM10*, and *ADAMTS1*). Pathway analysis of these data implicates the immune system and lipid metabolism, and for the first time tau binding proteins and APP metabolism. These findings show that genetic variants affecting APP and Aβ processing are not only associated with early-onset autosomal dominant AD but also with LOAD. Analysis of AD risk genes and pathways show enrichment for rare variants ($P = 1.32 \times 10^{-7}$) indicating that additional rare variants remain to be identified.

Main Text (word count: 1,991)

Previous work identified 19 genome-wide significant signals in addition to APOE[9], that influence risk for LOAD. These account for ~31% of the genetic variance of LOAD[2], leaving the majority of genetic risk uncharacterized[10]. To search for additional signals, we conducted a GWAS meta-analysis of non-Hispanic Whites (NHW) using a larger sample (17 new, 38 total datasets) from four consortia (ADGC, CHARGE, EADI, and GERAD). This sample increases the previous discovery sample (Stage 1) by 29% for cases and 13% for controls (N = 21,982 cases; 41,944 controls) (**Supplementary Table 1, 2** and **3**, and **Supplementary Note**). To sample both common and rare variants (minor allele frequency MAF ≥ 0.01, and MAF < 0.01, respectively), we imputed the discovery datasets using a 1000 Genomes reference panel consisting of 36,648,992 single-nucleotide variants, 1,380,736 insertions/deletions, and 13,805 structural variants. After quality control, 9,456,086 common variants and 2,024,664 rare variants were selected for analysis (a 62.7% increase from the 2013 common variant analysis). Genotype dosages were analyzed within each dataset, and then combined with meta-analysis (**Supplementary Figures 1 and 2** and **Supplementary Table 4**). The Stage 1 discovery meta-analysis was first followed by replication Stage 2 using the I-select chip we previously developed in Lambert et al (including

11,632 variants, n=19,884) and finally stage 3A (n=7,026). The final sample was 34,312 clinical AD cases and 57,943 controls.

Meta-analysis of Stages 1 and 2 produced 21 associations with $P \leq 5 \times 10^{-8}$ (**Table 1** and **Figure 1**). Of these, 18 were previously reported as genome-wide significant and three of them are signals not initially described in Lambert et al: the rare R47H *TREM2* coding variant previously reported by others[11–13]; *ECDH3* (rs7920721) which was recently identified as a potential genome-wide significant AD risk locus in several studies[23-25] and *ACE* (rs138190086). In addition, four signal showed suggestive association with a P-value<5.10^{-7} (respectively rs593742, rs830500, rsrs7295246 and rs138190086 for *ADAM10, ADAMTS1, ADAMTS20, and IQCK*).

Stage 3A replication and meta-analysis of all three stages for these 6 variants (excluding the *TREM2* signal, see **Supplementary Figure 1** for workflow) identified five genome-wide significant sites. In addition to ECDH3, this included four new genome-wide AD risk signals at *IQCK, ADAMTS1, ACE* and *ADAM10* not previously described in other AD GWAS (**Table 2, Supplementary Table X and Supplementary Figures 4-8**). *ACE* and *ADAM10* were previously reported as AD candidate genes[14–17] that were not replicated in some subsequent studies[18–22]. We also extended the analyses of the two loci (*NME8* and *MEF2C*) in stage 3 that were previously genome-wide significant in our 2013 meta-analysis. These loci were not significant in our current study and will deserve further investigations. Of note, GCTA-COJO[26] conditional analysis of the genome-wide loci indicates that *TREM2* and three other loci (*BIN1*, *ABCA7*, and *PTK2B/CLU*) have multiple independent LOAD association signals (**Supplementary Table 6**), suggesting that the genetic variance associated with some GWAS loci is probably under-estimated.

We also selected 23 SNPs from stage 1 (18 common variants in loci not well captured in the I-select chip, P<$5x10^{-6}$ and 5 rare variants with MAF<0.01, P<10^{-5}, see supplementary material and methods section for full selection criteria) for replication in stage 3B (including populations of stage 2 and stage 3A). We nominally replicated a relatively rare variant (rs71618613) within an intergenic region on 5p13.3 (MAF = 0.01; $P = 6.8x10^{-3}$; combined-P = $3.3x10^{-7}$)(**Table 2 and supplementary Table X**).

To evaluate the biological significance of the newly identified signals and those found previously, we pursued four strategies: expression-quantitative trait loci (eQTL) analyses, differential expression in AD versus control brains, gene cluster/pathway analyses, and expression in AD-relevant tissues[27,28]. For the 24 signals reported here, other evidence indicates that *APOE*[29,30], *ABCA7*[31,32], *BIN1*[33], *TREM2*[12,34], *SORL1*[35,36], *ADAM10*[37], *SPI1*[38], *and CR1*[39] are the true AD risk gene, though there is a possibility that multiple risk genes exist in these regions[40]. Because many GWAS loci are intergenic, and the closest gene to the sentinel variant may not be

the actual risk gene, in these analyses, we considered all genes within ±500kb of the sentinel variant linkage disequilibrium (LD) regions ($r^2 \geq 0.5$) for each locus as a candidate AD gene (**Supplementary Table 7**).

For eQTL analyses, we identified variants in LD with sentinel variants for each locus. For these variants, there were cis-acting eQTLs for 117 genes, with 92 eQTL-controlled genes in AD relevant tissues **(Supplementary Tables 8-11**). For our newly identified loci, the most significant eQTLs for the *ADAM10* signal were for *ADAM10* in blood ($P = 1.21 \times 10^{-13}$). For the *IQCK* signal, the top eQTL was for *DEF8* in monocytes ($P = 5.75 \times 10^{-48}$). For the ADAMTS1, signal, the most significant eQTL was for *ADAMTS1* in blood ($P = 7.56 \times 10^{-7}$). No eQTLs were found for the *ACE* locus. These results indicate that *ADAM10*, *ADAMTS1*, and *DEF8* were the genes responsible for the observed association signal. For previously identified loci, there were eQTLs for *BIN1* in monocytes ($P = 3.46 \times 10^{-67}$), *PVRIG* in blood at the *NYAP1* locus ($P = 2.02 \times 10^{-221}$), and *SLC24A4* in monocytes ($P = 1.27 \times 10^{-34}$).

For our differential expression studies of AD versus control brains, we used thirteen expression studies[41]. Of 469 protein coding genes within the genome-wide loci, we found 87 upregulated and 55 downregulated genes that were differentially expressed in the same direction in two or more studies. These include four genes at the *ADAM10* locus (*ADAM10* and *SLTM,* each upregulated in two studies; *AQP9*, downregulated in three studies; and *LIPC*, downregulated in two studies), three genes in the *IQCK* locus (*GPRC5B, CCP10*, and *GDE1* upregulated in 13, six and four studies, respectively), six genes in the *ACE* locus (*MAP3K3*, *KCNH6* and *FTSJ3*, downregulated in seven, two and two studies respectively; and *DDX42*, *PSMC5* and *TANC2*, downregulated in seven, five and three studies respectively), and three genes in the *ADAMTS1* locus (*ADAMTS1, CYYR1,* and *ADAMTS5,* upregulated in ten, two and two studies respectively) (**Supplementary Table 12**). For previously described loci, differentially expressed genes included *TFEB near TREM2, MS4A6A* (upregulated in 10 studies) at the chromosome 11 *MS4A* gene cluster, and *FERMT2* (upregulated in 9 studies) on chromosome 14, among others. Brain RNA-seq data reveals many of these differentially expressed candidate genes are expressed in AD-relevant cell types (**Supplementary Table 12**).

We conducted pathway analyses (MAGMA[42]) using five gene set resources. Analysis were conducted separately for common (MAF $\geq$ 0.01) and rare variants (MAF < 0.01). For common variants, we detected four function clusters including: 1) APP metabolism/Aβ-formation (regulation of beta-amyloid formation: $P = 4.56 \times 10^{-7}$ and regulation of amyloid precursor protein catabolic process: $P = 3.54 \times 10^{-6}$), 2) tau protein binding ($P = 3.19 \times 10^{-5}$), 3) lipid metabolism (four pathways including protein-lipid complex assembly: $P = 1.45 \times 10^{-7}$), and 4) immune response (P

= 6.32×10^{-5}) (**Table 3** and **Supplementary Table 13**). Enrichment of the four pathways remains after removal of genes in the *APOE* region. When APOE-region genes and genes in the vicinity of genome-wide significant genes are removed, tau shows moderate association and lipid metabolism and immune related pathways show strong associations (**Supplementary Table 14**). Genes driving these enrichments (i.e. having a gene-wide $P < 0.05$) include *SCNA*, a Parkinson's risk gene that may play a role in tauopathies[43,44], for the tau pathway, apolipoprotein genes (*APOM*, *APOA5*) and *ABCA1*, a major regulator of cellular cholesterol, for the lipid metabolism pathways, and 52 immune pathway genes (**Supplementary Table 15**). While no pathways were significantly enriched for rare variants, lipid and Aβ-pathways did have nominal significance in rare-variant-only analyses. Importantly, we also observe a highly significant correlation between common and rare pathway gene results ($P = 1.32 \times 10^{-7}$), suggesting that risk AD genes and pathways are enriched for rare variants. In fact, 50 different genes within tau, lipid, immunity and Aβ pathways show nominal association ($P < 0.05$) with LOAD (**Supplementary Table 15**).

To further explore the APP/Aβ-pathway enrichment we analyzed a comprehensive set of 335 APP metabolism genes[45] curated from the literature. We observed significant enrichment of this gene-set in common variants ($P = 2.27 \times 10^{-4}$; $P = 3.19 \times 10^{-4}$ excluding *APOE*), with both *ADAM10* and *ACE* nominally significant drivers of this result (**Table 4** and **Supplementary Table 16 and 17**). Several 'sub-pathways' were also significantly enriched in the common-variants including 'clearance and degradation of Aβ' and 'aggregation of Aβ', along with its subcategory 'microglia', the latter supporting the recent hypothesis that microglia play a large role in AD[13,46]. Nominal enrichment for risk from rare variants was found for the pathway 'aggregation of Aβ: chaperone' and 23 of the 335 genes.

To identify candidate genes for our novel loci, we combined results from eQTL, differential expression, AD-relevant tissue expression, and gene function/pathway analyses (**Table 5**). For our *ADAM10* signal, of the 17 genes within this locus, only *ADAM10* meets all our prioritization criteria. In addition, *ADAM10* is the most important α-secretase in the brain and a component of the non-amyloidogenic pathway of APP metabolism[47]. Over-expression of *ADAM10* in mouse models can halt Aβ production and subsequent aggregation[48]. Also two rare segregating familial LOAD *ADAM10* mutations increased Aβ plaque load in "Alzheimer-like" mice, with diminished α-secretase activity from the mutations likely the causal mechanism[16,37]. For the *IQCK* signal three of the 12 genes at the locus are potential candidate genes: *IQCK*, *DEF8*, and *GPRC5B.* The latter is a regulator of neurogenesis[49,50] and inflammatory signalling in obesity[51]. Of the 23 genes in the *ACE* locus, two meet three of the four prioritization criteria, *PSMC5*, a major regulator of major histocompatibility complex[52,53], and *CD79B,* a B lymphocyte antigen receptor sub-unit. Candidate

gene studies previously associate *ACE* variants with AD risk[17,54,55], including a strong association in the Wadi Ara, an Israeli Arab community with high risk of AD[21]. However, these studies yielded inconsistent results[18], and our work is the first to report a clear genome-wide association in NHW at this locus. While our analyses did not prioritize *ACE,* it should not be rejected as a candidate gene, as its expression in AD brain tissue is associated with Aβ load and AD severity[56]. Furthermore, CSF levels of the angiotensin-converting enzyme (Ace) are associated with Aβ levels[57] and LOAD risk[58], and studies show Ace can inhibit Aβ toxicity and aggregation[59] though it does not appear to regulate cerebral amyloidosis[60]. Another novel genome-wide locus reported here *ADAMTS1,* is within 665 kb of *APP* on chromosome 21. Of four genes at this locus (*ADAMTS1*, *ADAMTS5*, *CYYR1*, *CYYR1-AS1*) , our analyses nominates *ADAMTS1,* as the likely risk gene, though we cannot rule out that this signal is a regulatory element for *APP*. *ADAMTS1* is elevated in Down's Syndrome with neurodegeneration and AD[61] and is a potential neuroprotective gene[62,63,64], or a neuroinflammatory gene important to microglial response[65]. For previously reported loci, named for the closest gene, by applying the same approach for prioritization, our analysis highlights several genes as described in **Table 5**. It is also interesting to keep in mind that systematic biological screening have also highlighted some of these genes as involved in the APP metabolism (*FERMT2*) or Tau toxicity (*BIN1*, *CD2AP*, *FERMT2*, *CASS4*, *EPHA1*, *PTK2B*)[66–68].

Pathway, tissue and disease traits enrichment analysis supports the utility of our prioritization method, as the 68 prioritized genes are: 1) enriched in substantially more AD relevant pathways and processes, 2) enriched in AD relevant tissues such as monocytes (adjusted-*P*=1.75 x 10^{-6}) and macrophages (adjusted-*P*=6.46 x 10^{-3}), and 3) increased in associations of dementia-related traits (**Supplementary Table 18 and 19**).

Our work identifies four new genome-wide associations for LOAD, and shows that GWAS data combined with high-quality imputation panels can reveal rare disease risk variants (i.e. *TREM2*). The enrichment of rare-variants in pathways associated with AD indicates that additional rare-variants remain to be identified, and larger samples and better imputation panels will facilitate identifying these rare variants. While these rare-variants may not contribute substantially to the predictive value of genetic findings, it will add to the understanding of disease mechanism and potential drug targets. Discovery of the risk genes at genome-wide loci remains challenging, but we demonstrate that converging evidence from existing and new analyses can prioritize risk genes. We also show that APP metabolism is not only associated with early-onset but also late-onset AD, suggesting that therapies developed by studying early-onset families will apply to the more common late-onset form of the disease. Finally, our analysis showing tau is involved in late-

onset AD confirms that therapies targeting tangle formation/degradation could potentially affect late-onset AD.

References

1. Thies, W., Bleiler, L. & Alzheimer's, A. 2013 Alzheimer's disease facts and figures. *Alzheimers Dement* **9,** 208–245 (2013).
2. Ridge, P. G. *et al.* Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol. Aging* 1–8 (2016). doi:10.1016/j.neurobiolaging.2016.02.024
3. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303,** 1832–1840 (2010).
4. Jun, G. *et al.* Meta-analysis confirms CR1, CLU, and PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes. *Arch Neurol* **67,** 1473–1484 (2010).
5. Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43,** 436–441 (2011).
6. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41,** 1088–1093 (2009).
7. Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43,** 429–435 (2011).
8. Lambert, J. C. *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41,** 1094–1099 (2009).
9. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–8 (2013).
10. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* **63,** 168–174 (2006).
11. Guerreiro, E. *TREM2* and Neurodegenerative Disease. *N. Engl. J. Med.* **369,** 1564–1570 (2013).
12. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368,** 107–16 (2013).
13. Sims, R., van der Lee, S. J., Naj, A. C., Bellenguez, C. & et al. Novel rare coding variants in PLCG2, ABI3 and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Under Rev. Nat. Genet.* (2016).
14. Guo, Z. *et al.* Head injury and the risk of AD in the MIRAGE study. *Neurology* **54,** 1316–1323 (2000).
15. Vassar, R. ADAM10 prodomain mutations cause late-onset Alzheimer's disease: not just the latest FAD. *Neuron* **80,** 250–253 (2013).
16. Kim, M. *et al.* Potential late-onset Alzheimer's disease-associated mutations in the

ADAM10 gene attenuate {alpha}-secretase activity. *Hum Mol Genet* **18,** 3987–3996 (2009).

17. Kehoe, P. G. *et al.* Variation in DCP1, encoding ACE, is associated with susceptibility to Alzheimer disease. *Nat. Genet.* **21,** 71–72 (1999).
18. Wang, X.-B. *et al.* Angiotensin-Converting Enzyme Insertion/Deletion Polymorphism Is Not a Major Determining Factor in the Development of Sporadic Alzheimer Disease: Evidence from an Updated Meta-Analysis. *PLoS One* **9,** e111406 (2014).
19. Cai, G. *et al.* Evidence against a role for rare ADAM10 mutations in sporadic Alzheimer disease. *Neurobiol Aging* **33,** 416–417 e3 (2012).
20. Belbin, O. *et al.* A multi-center study of ACE and the risk of late-onset Alzheimer's disease. *J Alzheimers Dis* **24,** 587–597 (2011).
21. Meng, Y. *et al.* Association of polymorphisms in the Angiotensin-converting enzyme gene with Alzheimer disease in an Israeli Arab community. *Am J Hum Genet* **78,** 871–877 (2006).
22. Wakutani, Y. *et al.* Genetic analysis of vascular factors in Alzheimer's disease. *Ann. N. Y. Acad. Sci.* **977,** 232–8 (2002).
23. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case–control association mapping by proxy using family history of disease. *Nat. Genet.* **49,** 325–331 (2017).
24. Desikan, R. S. *et al.* Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. *Circulation* **131,** 2061–2069 (2015).
25. Jun, G. R. *et al.* Transethnic genome-wide scan identifies novel Alzheimer's disease loci. *Alzheimer's Dement.* 1–12 (2017). doi:10.1016/j.jalz.2016.12.012
26. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44,** 369–75, S1-3 (2012).
27. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34,** 11929–11947 (2014).
28. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89,** 37–53 (2016).
29. Corder, E. H. *et al.* Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat. Genet.* **7,** 180–184 (1994).
30. Kim, J., Basak, J. M. & Holtzman, D. M. The role of apolipoprotein E in Alzheimer's disease. *Neuron* **63,** 287–303 (2009).

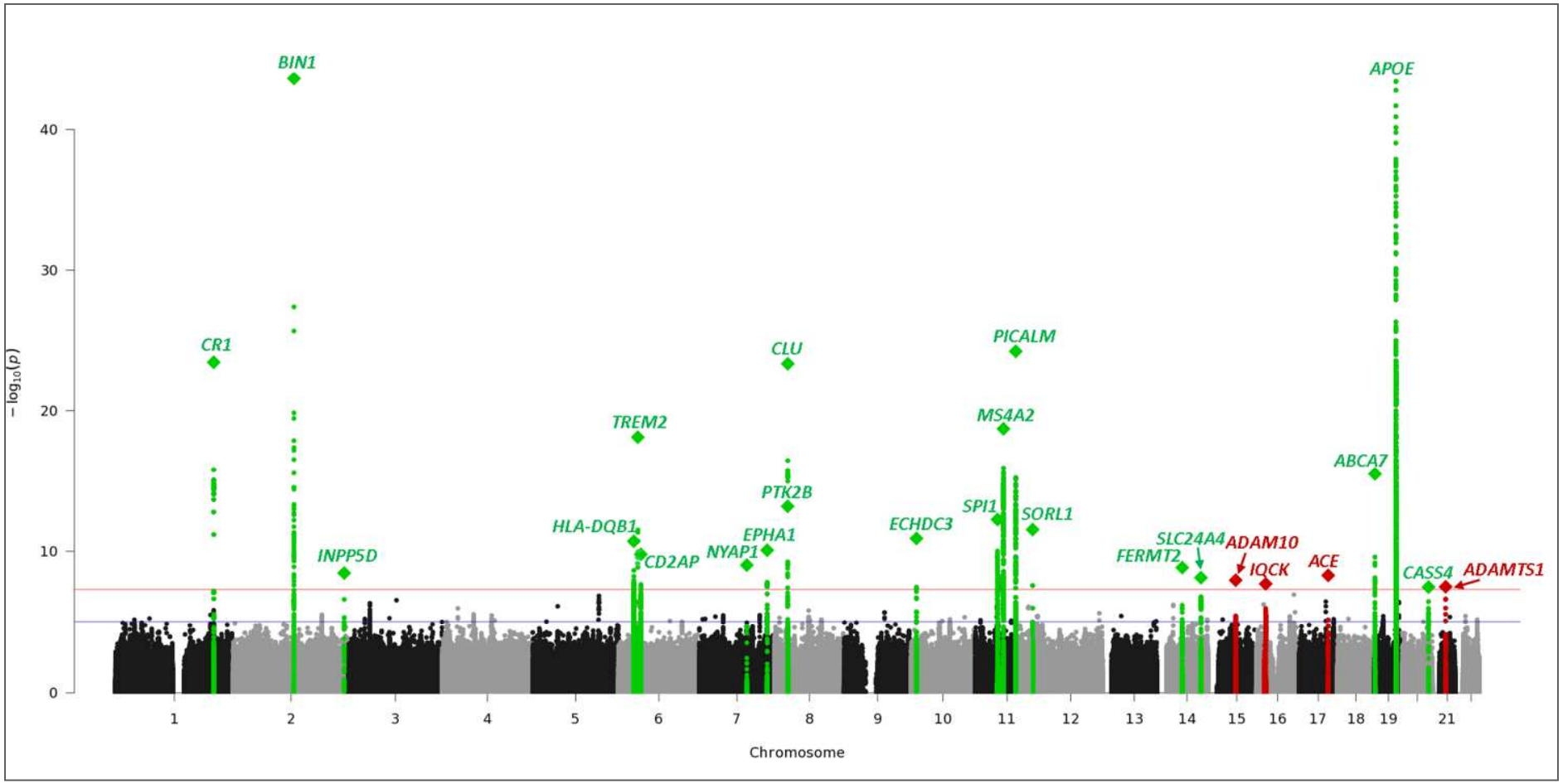
31. Steinberg, S. *et al.* Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat. Genet.* 26–29 (2015). doi:10.1038/ng.3246
32. Vasquez, J. B., Fardo, D. W. & Estus, S. ABCA7 expression is associated with Alzheimer's disease polymorphism and disease status. *Neurosci Lett* **556,** 58–62 (2013).
33. Chapuis, J. *et al.* Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol Psychiatry* (2013). doi:10.1038/mp.2013.1
34. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* **368,** 117–27 (2013).
35. Rogaeva, E. *et al.* The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet* **39,** 168–177 (2007).
36. Vardarajan, B. N. *et al.* Coding mutations in *SORL 1* and Alzheimer disease. *Ann. Neurol.* **77,** 215–227 (2015).
37. Suh, J. *et al.* ADAM10 missense mutations potentiate beta-amyloid accumulation by impairing prodomain chaperone function. *Neuron* **80,** 385–401 (2013).
38. Huang, K. *et al.* A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *bioRxiv* **20,** 1–9 (2017).
39. Brouwers, N. *et al.* Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol Psychiatry* **17,** 223–233 (2012).
40. Flister, M. J. *et al.* Identifying multiple causative genes at a single GWAS locus. *Genome Res* **23,** 1996–2002 (2013).
41. Bai, Z. *et al.* AlzBase: an Integrative Database for Gene Dysregulation in Alzheimer's Disease. *Mol. Neurobiol.* **53,** 310–319 (2016).
42. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11,** 1–19 (2015).
43. Stefanis, L. alpha-Synuclein in Parkinson's disease. *Cold Spring Harb. Perspect. Med.* **2,** 1–23 (2012).
44. Takeda, A. *et al.* C-terminal alpha-synuclein immunoreactivity in structures other than Lewy bodies in neurodegenerative disorders. *Acta Neuropathol.* **99,** 296–304 (2000).
45. Campion, D., Pottier, C., Nicolas, G., Le Guennec, K. & Rovelet-Lecrux, A. Alzheimer disease: modeling an Aβ-centered biological network. *Mol. Psychiatry* 1–11 (2016). doi:10.1038/mp.2016.38
46. Gandy, S. & Heppner, F. L. Microglia as dynamic and essential components of the amyloid hypothesis. *Neuron* **78,** 575–577 (2013).

47. Haass, C., Kaether, C., Thinakaran, G. & Sisodia, S. Trafficking and proteolytic processing of APP. *Cold Spring Harb Perspect Med* **2,** a006270 (2012).
48. Postina, R. *et al.* A disintegrin-metalloproteinase prevents amyloid plaque formation and hippocampal defects in an Alzheimer disease mouse model. *J. Clin. Invest.* **113,** 1456–1464 (2004).
49. Kurabayashi, N., Nguyen, M. D. & Sanada, K. The G protein-coupled receptor GPRC5B contributes to neurogenesis in the developing mouse neocortex. *Development* **140,** 4335–4346 (2013).
50. Cool, B. H. *et al.* A flanking gene problem leads to the discovery of a Gprc5b splice variant predominantly expressed in C57BL/6J mouse brain and in maturing neurons. *PLoS One* **5,** (2010).
51. Kim, Y.-J., Sano, T., Nabetani, T., Asano, Y. & Hirabayashi, Y. GPRC5B Activates Obesity-Associated Inflammatory Signaling in Adipocytes. *Sci. Signal.* **5,** ra85-ra85 (2012).
52. Bhat, K. *et al.* The 19S proteasome ATPase Sug1 plays a critical role in regulating MHC class II transcription. *Mol. Immunol.* **45,** 2214–24 (2008).
53. Inostroza-Nieves, Y., Venkatraman, P. & Zavala-Ruiz, Z. Role of Sug1, a 19S proteasome ATPase, in the transcription of MHC I and the atypical MHC II molecules, HLA-DM and HLA-DO. *Immunol. Lett.* **147,** 67–74 (2012).
54. Farrer, L. A. *et al.* Association Between Angiotensin-Converting Enzyme and Alzheimer Disease. *New Engl. J. Med.* **57,** 210–14 (2000).
55. Lehmann, D. J. *et al.* Large meta-analysis establishes the ACE insertion-deletion polymorphism as a marker of Alzheimer's disease. *Am. J. Epidemiol.* **162,** 305–17 (2005).
56. Miners, J. S. *et al.* Angiotensin-converting enzyme levels and activity in Alzheimer's disease: Differences in brain and CSF ACE and association with ACE1 genotypes. *Am. J. Transl. Res.* **1,** 163–177 (2009).
57. Jochemsen, H. M. *et al.* The association of angiotensin-converting enzyme with biomarkers for Alzheimer's disease. *Alzheimer's Res. Ther.* **6,** 1–10 (2014).
58. Kauwe, J. S. K. *et al.* Genome-Wide Association Study of CSF Levels of 59 Alzheimer's Disease Candidate Proteins: Significant Associations with Proteins Involved in Amyloid Processing and Inflammation. *PLoS Genet.* **10,** e1004758 (2014).
59. Baranello, R. J. *et al.* Amyloid-beta protein clearance and degradation (ABCD) pathways and their role in Alzheimer's disease. *Curr. Alzheimer Res.* **12,** 32–46 (2015).

60. Eckman, E. A., Watson, M., Marlow, L., Sambamurti, K. & Eckman, C. B. Alzheimer's disease beta-amyloid peptide is increased in mice deficient in endothelin-converting enzyme. *J. Biol. Chem.* **278,** 2081–2084 (2003).
61. Miguel, R. F., Pollak, A. & Lubec, G. Metalloproteinase ADAMTS-1 but not ADAMTS-5 is manifold overexpressed in neurodegenerative disorders as Down syndrome, Alzheimer's and Pick's disease. *Brain Res. Mol. Brain Res.* **133,** 1–5 (2005).
62. Suttkus, A. *et al.* Aggrecan, link protein and tenascin-R are essential components of the perineuronal net to protect neurons against iron-induced oxidative stress. *Cell Death Dis.* **5,** e1119 (2014).
63. Végh, M. J. *et al.* Reducing hippocampal extracellular matrix reverses early memory deficits in a mouse model of Alzheimer's disease. *Acta Neuropathol. Commun.* **2,** 76 (2014).
64. Morawski, M., Filippov, M., Tzinia, A., Tsilibary, E. & Vargova, L. ECM in brain aging and dementia. *Prog. Brain Res.* **214,** 207–27 (2014).
65. Wilcock, D. M. Neuroinflammation in the aging down syndrome brain; lessons from Alzheimer's disease. *Curr Gerontol Geriatr Res* **2012,** 170276 (2012).
66. Dourlen, P. *et al.* Functional screening of Alzheimer risk loci identifies PTK2B as an in vivo modulator and early marker of Tau pathology. *Mol. Psychiatry* **22,** 874–883 (2017).
67. Chapuis, J. *et al.* Genome-wide, high-content siRNA screening identifies the Alzheimer's genetic risk factor FERMT2 as a major modulator of APP metabolism. *Acta Neuropathol.* **133,** 955–966 (2017).
68. Shulman, J. M. *et al.* Functional screening in Drosophila identifies Alzheimer's disease susceptibility genes and implicates tau-mediated mechanisms. *Hum. Mol. Genet.* **23,** 870–877 (2014).

FIGURES

Figure 1. Manhattan plot of meta-analysis of Stage 1, 2 and 3 results for genome-wide association with Alzheimer's disease. The threshold for genome-wide significance ($P < 5 \times 10^{-8}$) is indicated by the red line, while the blue line represents the suggestive threshold ($P < 1 \times 10\text{-}5$). Loci previously identified by the Lambert et al. 2013 IGAP GWAS are shown in green, and newly associated loci are shown in red. Diamonds represent variants with the smallest *P* values for each genome-wide locus.



TABLES

Table 1. Summary of discovery stage 1, stage 2 and overall meta-analyses results for identified loci reaching genome-wide significance after stages 1 and 2.

						Stage 1 Discovery (n=65,773)		Stage 2 (n=19,884)		Overall Stages 1 + Stage 2 (n=85,657)	
SNP[a]	Chr.	Position[b]	Closest gene[c]	Major/ minor alleles	MAF[d]	OR (95% CI)[e]	Meta P value	OR (95% CI)[e]	Meta P value	OR (95% CI)[e]	Meta P value
Previous genome-wide significant loci still reaching significance											
rs4844610	1	207802552	*CR1*	C/A	0.187	1.16 (1.12-1.19)	8.2×10^{-16}	1.20 (1.13-1.27)	3.8×10^{-10}	1.17 (1.14-1.21)	3.6×10^{-24}
rs6733839	2	127892810	*BIN1*	C/T	0.406	1.18 (1.15-1.22)	4.0×10^{-28}	1.23 (1.18-1.29)	2.0×10^{-18}	1.20 (1.17-1.23)	2.5×10^{-44}
rs10933431	2	233981912	*INPP5D*	C/G	0.223	0.90 (0.87-0.94)	2.6×10^{-7}	0.92 (0.87-0.97)	3.2×10^{-3}	0.91 (0.88-0.94)	3.4×10^{-9}
rs78738018	6	32575406	*HLA-DQB1*	T/A	0.270	1.10 (1.06-1.14)	5.1×10^{-8}	1.11 (1.06-1.17)	5.6×10^{-5}	1.10 (1.07-1.19)	1.9×10^{-11}
rs75932628	6	41129252	*TREM2*	C/T	0.008	2.01 (1.82-2.21)	2.9×10^{-12}	2.37 (1.74-3.22)	4.3×10^{-8}	2.42 (1.99-2.95)	7.5×10^{-19}
rs9473117	6	47431284	*CD2AP*	A/C	0.280	1.09 (1.06-1.12)	2.3×10^{-7}	1.11 (1.05-1.16)	1.0×10^{-4}	1.09 (1.06-1.12)	1.6×10^{-10}
rs12539172	7	100091795	*NYAP1*[g]	C/T	0.331	0.93 (0.91-0.96)	2.1×10^{-5}	0.89 (0.84-0.93)	2.1×10^{-6}	0.92 (0.90-0.95)	9.3×10^{-10}
rs11762262	7	143107876	*EPHA1*	T/C	0.197	0.90 (0.87-0.94)	2.1×10^{-8}	0.91 (0.86-0.96)	9.6×10^{-4}	0.90 (0.88-0.93)	8.2×10^{-11}
rs73223431	8	27219987	*PTK2B*	C/T	0.367	1.10 (1.07-1.13)	8.3×10^{-10}	1.11 (1.06-1.16)	1.5×10^{-5}	1.10 (1.07-1.13)	6.3×10^{-14}
rs9331896	8	27467686	*CLU*	T/C	0.387	0.88 (0.85-0.91)	3.6×10^{-16}	0.87 (0.83-0.91)	1.7×10^{-9}	0.88 (0.85-0.90)	4.6×10^{-24}
rs3740688	11	47380340	*SPI1*[h]	T/G	0.448	0.91 (0.89-0.94)	9.7×10^{-11}	0.93 (0.88-0.97)	1.2×10^{-3}	0.92 (0.89-0.94)	5.4×10^{-13}
rs7933202	11	59936926	*MS4A2*	A/C	0.391	0.89 (0.86-0.92)	2.1×10^{-15}	0.90 (0.86-0.95)	1.6×10^{-5}	0.89 (0.87-0.92)	1.9×10^{-19}
rs3851179	11	85868640	*PICALM*	G/A	0.356	0.89 (0.86-0.91)	5.6×10^{-16}	0.85 (0.81-0.89)	6.1×10^{-11}	0.88 (0.86-0.90)	6.0×10^{-25}
rs11218343	11	121435587	*SORL1*	T/C	0.040	0.81 (0.76-0.88)	2.6×10^{-8}	0.77 (0.68-0.87)	1.2×10^{-5}	0.80 (0.75-0.85)	2.8×10^{-12}
rs17125924	14	53391680	*FERMT2*	T/C	0.093	1.13 (1.08-1.19)	6.6×10^{-7}	1.15 (1.06-1.25)	5.0×10^{-4}	1.14 (1.09-1.18)	1.4×10^{-9}
rs12881735	14	92932828	*SLC24A4*	T/C	0.221	0.92 (0.88-0.95)	4.9×10^{-7}	0.92 (0.87-0.97)	4.2×10^{-3}	0.92 (0.89-0.94)	7.3×10^{-9}
rs3752246	19	1056492	*ABCA7*	C/G	0.185	1.13 (1.09-1.18)	6.6×10^{-10}	1.18 (1.11-1.25)	4.7×10^{-8}	1.15 (1.11-1.18)	3.1×10^{-16}
rs429358	19	45411941	*APOE*	T/C	0.215	0.30 (0.28-0.31)	1.2×10^{-881}	*APOE* region not carried forward to replication stage			
rs6024870	20	54997568	*CASS4*	G/A	0.087	0.88 (0.83-0.93)	1.1×10^{-6}	0.90 (0.83-0.97)	9.0×10^{-3}	0.88 (0.85-0.92)	3.4×10^{-8}
New genome-wide significant loci reaching significance											
rs138190086	7	61538148	*ACE*	G/A	0.02	1.29 (1.15-1.44)	7.4×10^{-6}	1.41 (1.18-1.69)	1.8×10^{-4}	1.32 (1.20-1.45)	7.5×10^{-9}
rs7920721	10	11720308	*ECDH3*	A/G	0.389	1.08 (1.05-1.11)	1.9×10^{-7}	1.07 (1.02-1.12)	3.2×10^{-3}	1.08 (1.05-1.11)	2.3×10^{-9}
Previous genome-wide significant loci not reaching significance											
rs190982[f]	5	88223420	*MEF2C*	A/G	0.390	0.95 (0.92-0.97)	2.8×10^{-4}	0.93 (0.89-0.98)	2.7×10^{-3}	0.94 (0.92-0.97)	2.8×10^{-6}
rs4723711[f]	7	37844263	*NME8*	A/T	0.356	0.95 (0.92-0.98)	2.7×10^{-4}	0.91 (0.87-0.95)	9.5×10^{-5}	0.94 (0.92-0.96)	2.8×10^{-7}

[a]Variants showing the best level of association after meta-analysis of stages 1 and 2.
[b]Build 37, assembly hg19.
[c]Based on position of top SNP in reference to the refSeq assembly
[d]Average in the discovery sample.
[e]Calculated with respect to the minor allele.
[f]Not replicated in stage 2.
[g]Previously the *ZCWPW1* locus.
[h]Previously the *CELF1* locus.

Table 2. Summary of discovery Stage 1, Stage 2, Stage 3 (A and B), and overall meta-analyses results for potential novel loci reaching P <5.10^{-7}.

						Stage 1 Discovery (n=65,773)		Stage 2 (n=19,884)		Stage 3A (n=7,026)		Overall (n=91,683)	
SNP[a]	**Chr.**	**Position[b]**	**Closest gene[c]**	**Major/Minor allele**	**MAF[d]**	**OR (95% CI)[e]**	**Meta *P***	**OR (95% CI)[e]**	**Meta *P***	**OR (95% CI)[e]**	**Meta *P***	**OR (95% CI)[e]**	**Meta *P***
rs7920721[f]	10	11720308	*ECHDC3*	A/G	0.389	1.08 (1.05-1.11)	1.9 x 10^{-7}	1.07 (1.02-1.12)	3.2 x 10^{-3}	1.13 (1.06-1.22)	5.9 x 10^{-4}	1.08 (1.06-1.11)	1.2 x 10^{-11}
rs593742	15	59045774	*ADAM10*	A/G	0.295	0.94 (0.91-0.97)	3.0 x 10^{-5}	0.92 (0.87-0.97)	8.8 x 10^{-4}	0.91 (0.84-0.99)	2.5 x 10^{-2}	0.93 (0.91-0.95)	1.1 x 10^{-8}
rs7185636	16	19808163	*IQCK*	T/C	0.18	0.92 (0.89-0.96)	2.5 x 10^{-5}	0.90 (0.85-0.95)	7.6 x 10^{-4}	0.93 (0.85-1.01)	9.6 x 10^{-2}	0.92 (0.89-0.95)	2.0 x 10^{-8}
rs138190086	17	61538148	*ACE*	G/A	0.02	1.29 (1.15-1.44)	7.4 x 10^{-6}	1.41 (1.18-1.69)	1.8 x 10^{-4}	1.18 (0.90-1.55)	2.4 x 10^{-1}	1.31 (1.19-1.43)	5.0 x 10^{-9}
rs2830500	21	28156856	*ADAMTS1*	C/A	0.308	0.92 (0.89-0.95)	2.5 x 10^{-7}	0.95 (0.91-1.00)	5.7 x 10^{-2}	0.95 (0.88-1.02)	1.7 x 10^{-1}	0.93 (0.91-0.96)	3.2 x 10^{-8}
rs190982	*5*	*88223420*	*MEF2C*	*A/G*	*0.390*	*0.95 (0.92-0.97)*	*2.8 x 10^{-4}*	*0.93 (0.89-0.98)*	*2.7 x 10^{-3}*	*0.92 (0.86-0.99)*	*2.4 x 10^{-2}*	*0.94 (0.92-0.96)*	*2.4x10^{-7}*
rs4723711	*7*	*37844263*	*NME8*	*A/T*	*0.356*	*0.95 (0.92-0.98)*	*2.7 x 10^{-4}*	*0.91 (0.87-0.95)*	*9.5 x 10^{-5}*	*0.96 (0.90-1.03)*	*3.0 x 10^{-1}*	*0.94 (0.92-0.96)*	*2.0 x 10^{-7}*
						Stage 1 Discovery (n=65,773)				**Stage 3B (n=26,910)**		**Overall (n=91,683)**	
SNP[a]	**Chr.**	**Position[b]**	**Closest gene[c]**	**Major/Minor allele**	**MAF[d]**	**OR (95% CI)[e]**	**Meta *P***			**OR (95% CI)[e]**	**Meta *P***	**OR (95% CI)[e]**	**Meta *P***
rs71618613	5	29005878	*5p13.3*	A/C	0.01	0.68 (0.57-0.80)	9.8 x 10^{-6}	-	-	0.76 (0.63-0.93)	6.8 x 10^{-3}	0.71 (0.63-0.81)	3.3 x 10^{-7}

[a]SNPs showing the best level of association after meta-analysis of stages 1, 2 and 3.
[b]Build 37, assembly hg19.
[c]Based on position of top SNP in reference to the refSeq assembly
[d]Average in the discovery sample.
[e]Calculated with respect to the minor allele.
[f]Recently identified as a LOAD locus in two separate 2017 studies

Table 3. Significant pathways (q-value≤0.05) from MAGMA pathway analysis for common SNV and rare SNV subsets.

Pathway	N genes in pathway in dataset	Common SNVs *P**	Common SNVs q-value	Rare SNVs *P**	Rare SNVs q-value	Pathway description
GO:65005	20	1.45E-07*	9.53E-04	6.76E-02	8.42E-01	protein-lipid complex assembly
GO:1902003	10	4.56E-07*	1.49E-03	4.94E-02	8.42E-01	regulation of beta-amyloid formation
GO:32994	39	1.16E-06*	2.54E-03	1.78E-02	8.17E-01	protein-lipid complex
GO:1902991	12	3.54E-06*	5.80E-03	5.66E-02	8.42E-01	regulation of amyloid precursor protein catabolic process
GO:43691	17	5.55E-06*	6.75E-03	3.08E-02	8.17E-01	reverse cholesterol transport
GO:71825	35	6.18E-06*	6.75E-03	1.27E-01	8.42E-01	protein-lipid complex subunit organization
GO:34377	18	1.64E-05*	1.53E-02	1.82E-01	8.42E-01	plasma lipoprotein particle assembly
GO:48156	10	3.19E-05*	2.61E-02	7.77E-01	8.54E-01	tau protein binding
GO:2253	382	6.32E-05*	4.60E-02	2.09E-01	8.42E-01	activation of immune response

*Significant after FDR-correction (q-value≤0.05)

Table 4. Top results of pathway analysis of Aβ-beta centered biological network from Campion et al (see Supplementary Table 12 for full results).

Category	Subcategory	N Genes	Common SNVs P 0kb	Common SNVs P 35kb-10kb	Rare SNVs P 0kb	Rare SNVs P 35kb-10kb
Aβ -centered biological network (all genes)	--	331	2.27E-04*	1.54E-04*	8.26E-01	5.19E-01
Clearance and degradation of Aβ	--	74	2.18E-04*	3.27E-03	3.13E-01	5.11E-01
Clearance and degradation of Aβ	Microglia	47	2.24E-04*	1.83E-02	2.49E-01	6.87E-01
Aggregation of Aβ	--	35	7.09E-04*	9.93E-03	9.02E-02	1.68E-01
Aggregation of Aβ	Miscellaneous	21	1.08E-03*	3.38E-02	9.53E-02	1.90E-01
APP processing and trafficking	Clathrin/caveolin-dependent endocytosis	10	1.19E-03	1.15E-02	3.64E-01	1.84E-01
Mediator of Aβ toxicity	--	51	3.82E-02	4.69E-02	5.89E-01	5.70E-01
Mediator of Aβ toxicity	Calcium homeostasis	6	6.90E-02	1.21E-01	3.96E-01	2.54E-01
Mediator of Aβ toxicity	Miscellaneous	3	7.61E-02	2.35E-02	9.79E-01	7.61E-01
Clearance and degradation of Aβ	Enzymatic degradation of Aβ	15	7.77E-02	2.63E-02	6.10E-01	2.95E-01
Mediator of Aβ toxicity	Tau toxicity	20	9.03E-02	3.48E-01	7.17E-01	6.85E-01
Aggregation of Aβ	Chaperone	9	1.52E-01	3.09E-01	1.98E-01	1.13E-02

*Significant after Bonferroni correction for 33 pathway sets tested

Table 5. Top prioritized genes in significant loci based on biological evidence. Genes meeting at least 3 of 4 criteria in each locus are listed. The criteria include: 1) differential expression in at least one Alzheimer disease (AD) study, 2) expression in a tissue relevant to AD (astrocytes, neurons, microglia/macrophages, oligodendrocytes), 3) having an eQTL effect on the gene in any tissue, or having an eQTL on the gene in AD relevant tissue, and 4) being involved in a biological pathway enriched in AD (from the current study). Novel genome-wide loci from the current study are listed first, followed by known genome-wide loci.

Novel genome-wide loci							
Locus	**Number of genes in locus**	**Gene**	**Differential expression in AD**	**Expression in AD relevant tissue**	**eQTL in any tissue**	**eQTL in AD relevant tissue**	**In enriched pathway**
ADAM10	17	*ADAM10*					
IQCK	12	*GPRC5B*					
		IQCK					
		DEF8					
ACE	23	*PSMC5*					
		CD79B					
ADAMTS1	4	*ADAMTS1*					
Known genome-wide loci							
Locus	**Number of genes in locus**	**Gene**	**Differential expression in AD**	**Expression in AD relevant tissue**	**eQTL in any tissue**	**eQTL in AD relevant tissue**	**In enriched pathway**
CR1	13	*CD55*					
		CR1					
BIN1	10	*BIN1*					
INPP5D	14	*INPP5D*					
HLA-DQB1	59	*HLA-DPA1*					
		HLA-DRA					
		C4A					
		TNXB					
		PSMB9					
		HLA-DRB6					
		HLA-DRB1					
		HLA-DRB5					
		HLA-DQB1					
		AGPAT1					
		AGER					
		HLA-DQA1					
		C2					
		BRD2					
		HLA-DQB2					
		MICB					
TREM2	26	*TREM2*					
CD2AP	8	*CD2AP*					
NYAP1	60	*GAL3ST4*					
		EPHB4					
		PILRB					
		NYAP1					
		AGFG2					

		PILRA					
		GATS					
EPHA1	27	No gene meets 3 of the 4 criteria; 4 genes meet 2 of the 4 criteria					
PTK2B	12	*PTK2B*					
		CLU					
		SCARA3					
CLU	16	*CLU*					
ECHDC3	10	No gene meets 3 of the 4 criteria; 6 genes meet 2 of the 4 criteria					
SPI1	25	*PSMC3*					
		MTCH2					
		MADD					
		NUP160					
		PTPMT1					
		CELF1					
		RAPSN					
		NR1H3					
MS4A6A	24	*MS4A6A*					
		MS4A4A					
		OSBP					
PICALM	12	*SYTL2*					
		PICALM					
SORL1	4	*SORL1*					
FERMT2	10	*FERMT2*					
		PSMC6					
		STYX					
SLC24A4	11	*LGMN*					
		RIN3					
		SLC24A4					
ABCA7	49	*POLR2E*					
		STK11					
		CNN2					
		HMHA1					
		CFD					
		ABCA7					
		BSG					
CASS4	12	*CSTF1*					