Influence of coding variability in APP-Aβ metabolism genes in sporadic Alzheimer’s disease


*The Alzheimer’s Research UK (ARUK) Consortium: Peter Passmore, David Craig, Janet Johnston, Bernadette McGuinness, Stephen Todd, Queen’s University Belfast, UK; Reinhard Heun, Royal Derby Hospital, UK; Heike Kölsch, University of Bonn, Germany; Patrick G. Kehoe, University of Bristol, UK; Nigel M. Hooper, University of Leeds, UK; Emma R.L.C. Vardy, University of Newcastle, UK; David M. Mann, Stuart Pickering-Brown, University of Manchester, UK; Kristelle Brown, James Lowe, Kevin Morgan, University of Nottingham, UK; A. David Smith, Gordon Wilcock, Donald Warden, University of Oxford (OPTIMA), UK; Clive Holmes, University of Southampton, UK.

1. ABSTRACT

The cerebral deposition of Aβ42, a neurotoxic proteolitic derivate of amyloid precursor protein (APP), is a central event in Alzheimer’s disease (AD)(Amyloid hypothesis). Given the key role of APP-Aβ metabolism in AD pathogenesis, we selected 29 genes involved in APP processing, Aβ degradation and clearance. We then used exome and genome sequencing to investigate the single independent (single-variant association test) and cumulative (gene-based association test) effect of coding variants in these genes as potential susceptibility factors for AD, in a cohort composed of 435 sporadic and mainly late-onset AD cases and 801 elderly controls from North America and the UK. Our study shows that common coding variability in these genes does not play a major role for the disease development. In the single-variant association analysis, the main hits, which were nominally significant, were found to be very rare coding variants (MAF 0.3%-0.8%) that map to genes involved in APP processing (MEP1B), trafficking and recycling (SORL1), Aβ extracellular degradation (ACE) and clearance (LRP1). Moreover, four genes (ECE1, LYZ, TTR and MME) have been found as nominally associated to AD using c-alpha and SKAT tests. We suggest that Aβ degradation and clearance, rather than Aβ production, may play a crucial role in the etiology of sporadic AD.
Key words
Alzheimer’s disease, APP-Ãµ metabolism, exome sequencing, genome sequencing

2. INTRODUCTION

The cerebral deposition of Aß_{42} aggregates, insoluble neurotoxic derived of amyloid precursor protein (APP), is likely caused by an imbalance between Aß production and clearance and represents a key event in Alzheimer’s disease (Amyloid hypothesis) (Hardy and Selkoe, 2002). A growing body of evidence has pointed to the crucial role of APP-Aß metabolism in AD pathogenesis. First, the discovery of APP, PSEN1 and PSEN2 mutations showed that familial Alzheimer’s disease is linked to Aß_{42} overproduction (Goate et al., 1991)(Sherrington et al., 1995)(Rogaev et al., 1995). Second, genome-wide association studies (GWASs) identified several susceptibility loci associated with AD (APOE, BIN1, PICALM, CD33, ABCA7, CLU, MS4A6A, EPHA1, CR1, CD2AP, SORL1, CASS4) and regulating APP-Aß levels (Harold et al., 2009)(Lambert et al., 2009)(Seshadri et al., 2010)(Hollingworth et al., 2011)(Naj et al., 2011)(Kamboh et al., 2012)(Reitz et al., 2013)(Lambert et al., 2013). Third, next generation sequencing laid the ground for the discovery of TREM2 risk variants, highlighting the possible role of microglia in Aß clearance (Guerreiro et al., 2013)(Jonsson et al., 2013). Finally, although recent studies have shown that rare coding variability in PSEN1 may influence the susceptibility for late-onset apparently sporadic AD (Benitez et al., 2013) (Sassi et al., 2014), increases in Aß production, currently explain a minority of AD cases. By contrast, it is very likely that the majority of AD cases are caused by impaired degradation and clearance of Aß, which is produced at normal levels throughout life (Miners et al., 2008)(Selkoe, 2001). Despite the importance of APP-Aß metabolism in AD, the role of genes taking part in Aß production and catabolism as susceptibility factors for AD is still elusive and has not been extensively investigated. Therefore, in this study, we selected 29 genes known to be involved in APP and Aß processing: ADAM9, ADAM10, ADAM17, MEP1B, BACE1, BACE2, NCSTN, PEN2, APH1B, LRRTM3, APLP1, APBA1, SORL1, TTR, GPR3, ECE1, ECE2, IDE, CST3, CTSB, CTSD, LYZ, MME, ACE, MMP3, A2M, PLAT, KLK6 and LRP1.
We then analyzed the single and independent and the cumulative effect of protein coding variants in these genes from exome and genome sequencing data, in a cohort composed of 435 sporadic and mainly late-onset AD cases and 801 elderly controls from North America and UK.

3. MATERIALS AND METHODS

We used exome and genome sequencing data to identify common, low frequency, and rare coding variants in 29 genes involved in: Aβ production (ADAM9, ADAM10, ADAM17, BACE1, BACE2, NCSTN, PEN2, APH1B, MEP1B, LRRTM3, GPR3), APP stabilization (APLP1, APBA1), APP recycling (SORL1), Aβ deposition (TTR), intracellular degradation (ECE1, ECE2, IDE, CST3, CTSB, CTSD, LYZ, MME), extracellular degradation and clearance (ACE, MMP3, A2M, PLAT, KLK6, LRP1). These genes were chosen on the basis of PubMed based literature search and/or based on predicted protein interactions using STRING (http://string.embl.de/).

The cohort was composed of 435 sporadic AD cases and 801 elderly controls, neuropathologically and clinically confirmed, originating from the UK and North America. Among the cases, 127 (29.19%) were early-onset (<65 years of age) and 308 (70.8%) were late-onset (≥ 65 years of age). The mean age at disease onset was 67 years (range 36-94 years) for cases and the mean age of ascertainment was 79 years (range 60-103 years) for controls (Table 1).

The study was approved by the appropriate institutional review boards.

Exome sequencing

We performed whole exome sequencing on a discovery cohort of 435 sporadic and mainly late-onset AD cases and 590 elderly controls. DNA was extracted from blood or brain for cases and brain only for controls using standard protocols. Library preparation for next generation sequencing used DNA (between 1 µg and 3 µg) fragmented in a Covaris E210 (Covaris Inc.). Following fragmentation, DNA was end-repaired by 5’phosphorylation, using the Klenow polymerase. A poly-adenine tail was added to the 3’end of the phosphorylated fragment and ligated to Illumina adapters. After purification using an AMPure DNA Purification kit (Beckman Coulter, Inc), adapter-ligated products were amplified. The DNA library was then hybridized to
an exome capture library (NimbleGen SeqCap EZ Exome v2.0, Roche Nimblegen Inc. or TruSeq, Illumina Inc.) and precipitated using streptavidin-coated magnetic beads (Dynal Magnetic Beads, Invitrogen). Exome-enriched libraries were PCR-amplified, and then DNA hybridized to paired-end flow cells using a cBot (Illumina, Inc.) cluster generation system. Samples were sequenced on the Illumina HiSeq™ 2000 using 2x100 paired end reads cycles.

**Whole Genome sequencing**

Genome sequencing was performed in 211 elderly, clinically healthy controls, from the Cache County Study on Memory in Aging. All samples were sequenced with the use of Illumina HiSeq technology. Alignment was performed with the use of CASAVA software20 and variant calling was performed with the use of SAMtools (Li et al., 2009) and GATK (McKenna et al., 2010). This sequencing and variant calling was performed by our collaborators at Brigham Young University.

**Bioinformatic**

Sequence alignment and variant calling was performed against the reference human genome (UCSC hg19). Paired end sequence reads (2x100bp paired end read cycles) were aligned using the Burrows-Wheeler aligner (Li and Durbin, 2009). Format conversion and indexing were performed with Picard (www.picard.sourceforge.net/index.shtml). The Genome Analysis Toolkit (GATK) was used to recalibrate base quality scores, perform local re-alignments around indels and to call and filter the variants (McKenna et al., 2010). VCFtools was used to annotate gene information for the remaining novel variants. We used ANNOVAR software to annotate the variants (Wang et al., 2010). Variants were checked against established databases (1000 Genomes Project and dbSNP v.134). The protein coding effects of variants was predicted using SIFT, Polyphen2 and SeattleSeq Annotation (gvs.gs.washington.edu/SeattleSeqAnnotation). All variants within the coding regions of 29 candidate genes (A2M [NM_000014], ACE [NM_001178057], ADAM9 [NM_003816], ADAM10 [NM_001110], ADAM17 [NM_003183], APBA1 [NM_00116], APH1B
[NM_031301], APLP1 [NM_005166], BACE1 [NM_012104], BACE2 [NM_138992], CST3 [NM_000099], CTSB [NM_147781], CTSD [NM_001909], ECE1 [NM_001397], ECE2 [NM_032331], GPR3 [NM_005281], IDE [NM_004969], LRP1 [NM_002332], KLK6 [NM_002774], LRRTM3 [NM_178011], LYZ [NM_000239], MEP1B [NM_00592], MME [NM_007289], MMP3 [NM_002422], NCSTN [NM_015331], PLAT [NM_033011], PSENEN [NM_172341], SORL1 [NM_003105], TTR [NM_000371]) have been collected and analyzed. (Table S1) (See supplementary materials for details)

**Statistical Analysis**

For each variant, allele frequencies were calculated in cases and controls and Fisher’s exact test on allelic association was performed. All computations were performed in R (version x64 3.0.2, http://www.r-project.org/).

In this study, we have sufficient power (≥80%) to detect common SNVs (MAF > 5%) with modest effect (OR = 2) through single-variant association analysis; however, we had limited power (<80%) to detect very rare SNVs (MAF < 0.1%), even those with strong effect (OR > 4).

A p-value of 0.05 was set as a nominal significance threshold. Based on simple Bonferroni correction for multiple testing, the thresholds for single variant and gene-based association are defined by p-value = 1.1e^-4 (0.05/449 coding variants) and 1.7e^-3 (0.05/29 genes).

C-alpha test and SKAT are closely related, being both non-burden test, analyzing and collapsing the effect of genetic variants of different frequency (common and rare), effect (protective, damaging and neutral) and effect size (modest, moderate, strong). SKAT can be considered an expansion of the c-alpha test because overcomes some of its limits. Indeed, SKAT 1) can be applied also to the study of continuous traits; 2) do not need any permutation; 3) apply covariates to the study.

Linkage disequilibrium and haplotype block structure analyses were performed with Haploview v. 4.2 (http://www.broadinstitute.org/haploview).
4. RESULTS

The study population consisted of a total of 435 sporadic and mainly late-onset AD cases and 801 elderly controls of British and North American ancestry (Table 1).

We performed a single-variant and a single-gene association analysis in a pre-defined set of genes involved in APP processing (ADAM9, ADAM10, ADAM17, MEP1B, BACE1, BACE2, NCSTN, PEN2, APH1B, APLP1, APBA1), Aβ metabolism and catabolism (LRRTM3, LRP1, TTR, GPR3, SORL1, ECE1, ECE2, IDE, CST3, CTSB, CTSD, LYZ, MME, ACE, MMP3, A2M, PLAT, KLK6), including 68 Megabase pairs (Mbs) of coding sequence.

A total of 1,787 single nucleotide variants (SNVs) has been identified. Among these, 449 (25.12%) were nonsynonymous, 452 (25.29%) were synonymous and 886 (48.46%) UTR variants. Among the missense variants, 345 (76.83%) were rare (MAF<1%), 21 (4.67%) were low frequency (1%<MAF<5%) and 12 (2.67%) were common (MAF>5%). In addition, we report 71 novel coding variants (ExAC, released 13 January 2015, or dbSNP 137)(Table S2).

Only 28 coding variants (6.2%) cluster within haplotypes and could have been potentially tagged by GWASs and chip based fine-mapping approaches.

263 variants (58.6%) were described as damaging variants by at least 2 out of 3 in silico prediction softwares (SIFT, Polyphen and Mutation Taster). Importantly, genes involved in Aβ degradation and clearance harbor the highest relative frequency of coding and damaging variants. By contrast, genes taking part in APP processing and Aβ production, present the lowest relative frequency of coding and damaging variants, suggesting a higher degree of conservation of this last cluster of genes (Table S3 and S4).

Single coding variant association test

The single variants association test identified 6 nominally significant variants, clustering in genes involved in Aβ degradation and clearance (SORL1, MEP1B, ACE, CTSB and LRP1). Importantly, all
of these are very rare coding variants (0.3<MAF<0.8%) with moderate to strong effect size (3<OR<9.2). In addition, all of these except MEP1B p.R132Q and LRP1 p.R2864H have been predicted as damaging by 3 in silico software (SIFT, Polyphen2 and Mutation Taster), cluster in conserved domains (PhastCons score 0.9-1) and are highly expressed in the brain (http://biogps.org/) (Table 2).

For all these variants the minor allele was substantially more frequent in cases compared to controls, suggesting a possible role as a risk factor for AD. However, none of these reached statistically significance based on a corrected p-value (p-value <1.1e^-4). The study possessed relatively low power to detect any significant association between cases and controls for low frequency and rare variants. However, we analyzed these variants because we could not preclude the possibility that high effect risk alleles were present.

Gene-based association test

In addition to single-marker analysis, we performed gene-wide analysis to combine the joint signal from multiple variants within a gene and to provide greater statistical power than that for single-marker tests. All the variants (nonsynonymous, synonymous, UTRs) located within the studied genes and their exon-intron flanking regions were collapsed together and their joint effect has been studied. Genes involved in Aβ degradation and clearance were enriched for the lowest p-values. The combined effect of variants in ECE1 and LYZ and in TTR reached the nominal significance in the c-alpha and SKAT test, respectively (Table 3 and 4). There was a partial overlap between genes identified in the single-marker analysis and those with SKAT and c-alpha test.

5. DISCUSSION

The Amyloid cascade hypothesis is the main accepted hypothesis underlying AD pathology. Several genes within the APP-Aβ metabolism pathway have been reported as potential candidate genes for AD. However, coding variability among these has not been extensively investigated. The vast majority of reported studies are based on candidate gene approaches using array-based
SNP genotyping and are focused mainly on genes involved in Aβ catabolism (http://www.alzgene.org/). Thus, leaving low frequency and rare coding variants and genes involved in Aβ production largely unexplored.

GWASs and chip-based candidate gene approaches have shown that common and generally non-coding variability within these genes do not play a crucial role for AD development. The only exceptions to this general rule are represented by SORL1 and ABCA7, which have been reported associated with late-onset apparently sporadic and familial AD both with GWASs, candidate gene approaches and exome sequencing (Rogaeva et al., 2007)(Pottier et al., 2012) (Liu et al., 2014)(Steinberg et al., 2015).

In this study, we report a screening of genes known to be involved in the APP-Aβ metabolism (APP processing, Aβ production, degradation and clearance). We applied single-marker and gene-based association analyses, to investigate the independent and joint effect of coding coding variability within these genes in a cohort composed of 435 apparently sporadic and mainly late-onset AD cases and 801 elderly controls from North America and the UK.

In our cohort, genes involved in Aβ degradation and clearance harbour the highest relative frequency of rare and predicted damaging variants. Conversely, genes encoding for proteins playing important roles in Aβ production (α, β and γ secretases complex [ADAM9, ADAM10, ADAM17, BACE1, BACE2, NCSTN, PSENEN, APH1B]) presented the lowest relative frequency of rare coding variants (Table S3 and S4), suggesting a higher degree of conservation.

In the single-variant association analysis, the main hits were very rare coding variants (MAF 0.3%-0.8%) with strong effect sizes (3<OR<9.2) mapping to genes involved in APP cleavage (MEP1B), APP trafficking and recycling (SORL1), Aβ extracellular degradation (ACE) and clearance (LRP1). Our study was underpowered for the detection of rare and low frequency variants and these variants were nominally significant after Bonferroni correction. However, given the fact that 1) all these variants except MEP1B (R132Q) and LRP1 (R2864H) have been reported as damaging by 3 in silico prediction softwares; 2) cluster in well conserved domains (PhastCons score 0.9-1) and 3) all these genes are highly expressed in the brain (http://biogps.org/) we suggest that these variants may be potential functional and warrant a follow up in an extended sample size.
**SORL1** harbors the most significant variant identified (p. N2174T). This gene encodes for Sortilin related receptor (SORLA, also known as LR11), which mediates several intracellular sorting and trafficking functions through a VPS10 (vacuolar protein sorting protein 10) domain (Yamazaki et al., 1996). SORLA is highly expressed in the brain and it binds to APOE and APP. Multiple lines of evidence suggested **SORL1** as an excellent positional and functional candidate for AD (Feulner et al., 2010) (Lambert et al., 2013)(Rogaeva et al., 2007)(Miyashita et al., 2013)(Tan et al., 2009)(Lee et al., 2007) (Pottier et al., 2012).

Moreover, SORLA modulates APP recycling through the retromer complex, thereby influencing levels of Aβ (Andersen et al., 2005)(Offe et al., 2006).

**MEP1B** harbors the two main hits detected in our study (Q496K a R132Q). These are not in linkage disequilibrium (LD), thus suggesting 2 independent signals. **MEP1B** encodes for meprin ß, a zinc metalloprotease which may be involved in APP processing either taking part to the amyloidogenic pathway, cleaving APP in a ß-secretase manner, or to the constitutive pathway, through the activation of ADAM10 (Bien et al., 2012)(Jefferson et al., 2011)(Jefferson et al., 2013). In either way, *in vivo* and *in vitro* experiments associated **MEP1B** activation with increased Aβ levels, suggesting a pathogenic role in AD. **MEP1B** is expressed in the CNS, particularly high levels are found in the hippocampus.

The third strongest signal maps to **ACE** (p.T342M), encoding angiotensin I converting enzyme 1 (ACE1), a zinc metalloprotease, which regulates blood pressure. Multiple lines of evidence have shown that ACE1 can convert Aβ42 to Aβ40, a more soluble isoform and its activity is increased in AD brain, in proportion to the parenchymal Aβ load (Zou et al., 2009)(Hu et al., 2001)(Hemming and Selkoe, 2005).

Moreover, catepsin B, encoded by **CTSB** is a lysosomal cysteine proteinase that is highly expressed in neurons and microglial cells. *In vitro* and *in vivo* experiments have shown Aβ reduction or elevation by enhancing or inhibiting **CTSB**, respectively (Wang et al., 2012).

Finally, rs143285614 (p.R2864H) maps to the low-density lipoprotein receptor related protein 1 (LRP1) gene. A growing body of evidence has shown that LRP1 is a major Aβ clearance receptor in cerebral vascular smooth muscle cells and disturbance of this pathway contributes to Aβ
accumulation in the brain (Kanekiyo et al., 2012). In addition, LRP1 locus was identified as an AD candidate locus in consanguineous Israeli-Arab community (Farrer et al., 2003).

The c-alpha and SKAT tests reported ECE1 and LYZ and TTR, respectively, as main genes associated with sporadic and mainly LOAD, although after the correction, the association was only nominally significant. Very interestingly, the main genes identified with the single-marker and gene-based analysis play a pivotal role in the cardiovascular system and have been already signaled as potential risk factors for cerebral amyloid angiopathy (CAA) or vascular dementia (Revesz et al., 2009). First, ECE1 and ACE are crucial components of the renin-angiotensin cascade, that controls vascular pressure (Rossi et al., 1999). Second, ECE1, ACE and TTR are mostly expressed by endothelial cells in the CNS (http://web.stanford.edu/group/barres_lab/brain_rnaseq.html). In concert with previous studies, these results suggest a potential overlapping biology with shared risk factors between CAA, vascular dementia and AD.

In summary, our study shows that 1) common coding variability within genes involved in Aβ metabolism is not likely to play a critical role for AD development; 2) genes involved in Aβ production are more conserved than genes playing a crucial role in Aβ degradation and clearance, therefore less frequently involved in sporadic AD; 2) SORL1, MEP1B, CTSB, ACE and LRP1 harbour rare coding variants with strong effect size and likely to be functional and warrant further investigation in a bigger sample size; 3) ECE1, LYZ and TTR play a critical role in the cardiovascular system and are likely associated to AD risk. Finally, post-transcriptional and post-translational modifications, different expression patterns and epigenetic factors should be considered as possible mechanisms controlling the pathogenic role of genes involved in Aβ metabolism and therefore further investigated.

ACKNOWLEDGEMENTS

This study was supported by the Alzheimer's Research UK, the Medical Research Council (MRC), the Wellcome Trust/MRC Joint Call in Neurodegeneration Award (WT089698) to the UK Parkinson’s Disease Consortium (whose members are from the University College London
Institute of Neurology, the University of Sheffield, and the MRC Protein Phosphorylation Unit at the University of Dundee), grants (P50 AG016574, U01 AG006786, and R01 AG18023), the National Institute for Health Research Biomedical Research Unit in Dementia at University College London Hospitals, University College London; an anonymous donor, the Big Lottery (to Dr. Morgan); a fellowship from Alzheimer's Research UK (to Dr. Guerreiro); and the Intramural Research Programs of the National Institute on Aging and the National Institute of Neurological Disease and Stroke, National Institutes of Health (Department of Health and Human Services Project number, ZO1 AG000950-10). The MRC London Neurodegenerative Diseases Brain Bank and the Manchester Brain Bank from Brains for Dementia Research are jointly funded from ARUK and AS.

REFERENCES


