bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TECPR2 a positive regulator of autophagy is implicated in healthy brain ageing

John Alexander[1], Thomas Ströbel[2], Marianthi Georgitsi[1], Michael Schuster[3], Thomas Penz[3] Christoph Bock[3-5], Selma Hönigschnabl[6], Angelika Reiner[6], Peter Fischer[7], Peristera Paschou[1,8*], Gabor G. Kovacs[2*]

1: Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupoli, Greece;

2: Institute of Neurology, Medical University Vienna, Vienna, Austria;

3: CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria;

4: Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria;

5: Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany;

6: Institute of Pathology, Danube Hospital Vienna, Austria;

7: Psychiatric Department, Medical Research Society Vienna, D.C., Danube Hospital, Vienna, Austria;

8: Department of Biological Sciences, Purdue University, West Lafayette, USA

Corresponding authors:

Gabor G. Kovacs MD PhD; Institute of Neurology, AKH 4J, Währinger Gürtel 18-20, A-1097 Vienna, Austria; Phone: +43-1-40400-55070; Fax +43-1-40400-55110;

Email: gabor.kovacs@meduniwien.ac.at

and

Peristera Paschou PhD; Dept of Biological Sciences, Purdue University, 915 W State Str, West Lafayette, IN 47907, USA; Email ppaschou@purdue.edu

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Understanding the healthy brain aging process is key to uncovering the mechanisms leading to pathological age-related neurodegeneration, including progression to Alzheimer's disease (AD). Here, we report the first deep whole genome sequencing study aiming to identify variants that are associated specifically to healthy brain aging defined on both clinical and neuropathological level, thus tacking the issue of pathological heterogeneity that often underlies a clinical AD diagnosis. We studied samples from the VITA brain bank and followed an extreme phenotypic ends study design comparing neuropathologically "healthy" aging individuals above 80 years of age with pure AD patients of the same age. Focusing on the extreme ends of the phenotypic distribution, and potentially functional variants, we discover a single variant (*rs10149146)* carried by 53.6% of the "healthy" brain elderly individuals in our study (15/28 individuals) and none of the 12 AD cases. This variant lies on the autophagy and cell cycle associated *TECPR2* gene. Autophagy dysfunction has been previously implicated in multiple progressive neurodegenerative diseases. An additional non-synonymous variant on the *CINP* gene (encoding a cell-cycle checkpoint protein) is also found in 46% of healthy controls and absent from all the AD cases. *TECPR2* and *CINP* appear to be "partner" genes in terms of regulation and their associated transcription factors have been previously implicated in AD and neurodegeneration. Our study is the first to support the hypothesis that a *TECPR2* non-synonymous variant carries a significant neuroprotective effect pointing to key molecules for the involvement of autophagy and cell cycle control in protection from neurodegeneration.

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

As life expectancy increases and fertility rates fall, the world population is aging[1]. However, despite the longer periods of good health and extended periods of social engagement and productivity, aging may also be associated with more illness and dependency. One of the consequences of longer life expectancies is the increase in people with dementia and especially Alzheimer's disease (AD). The global projection of people living with dementia by 2050 is 115.4 millions[2], posing a major burden on society and rendering imperative the need for understanding disease etiology and developing prevention strategies.

AD and other late onset forms of common neurodegenerative disorders are multifactorial and heterogeneous in nature and their genetic background remains poorly understood. Apart from their complex etiology (multiple genetic susceptibility factors), interaction with lifestyle and environmental factors leads to the onset of symptoms. Considering AD and including the *APOE* ε4 haplotype, approximately 61% of the population attributable risk of late onset AD has been explained[3]. A recent meta-analysis of 74,046 individuals confirmed eight and reported 11 newly associated AD susceptibility loci[4–7] A gene-based approach of the same dataset uncovered two more susceptibility loci[8]. However, the abovementioned genome-wide association studies implicate variants that account only for a small proportion of the estimated heritability of AD, leaving the substantial rest of the proportion unidentified. Attention has been turned to rare variants present in less than 1% of the population under the "common disease-multiple rare variants" hypothesis, currently amenable to large-scale analysis via next generation sequencing technologies[9]. In fact, low-frequency missense variants have been found to confer either strong protection or elevated risk of AD and cognitive decline[10–12].

On the other hand, longevity and healthy aging show a heritability of 20-50%[13,14]. One might anticipate that genetic factors which increase the risk of common complex neurodegenerative diseases such as AD would negatively affect life-span and be less common among long-lived people as compared to younger aged individuals[15]. Conversely, the genetic contribution to longevity is hypothesized to be greatest at the oldest ages[16]; however, not much is known regarding how genetic factors might influence healthy brain aging and cognition maintenance[17].The *TOMM40-APOE* region on chromosome 19q13.32 remains the most significantly associated locus with both longevity and cognition[16,18].

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To date, most genetic studies are based on a definition of case status using only clinical criteria. However, presence of multiple pathologies (i.e., combined neurodegenerative and vascular disorders) in the elderly with cognitive decline is rather the rule than the exception[19,20]. Therefore, pure AD pathology in demented individuals above 80 years of age is less frequent[19]. Moreover, individuals, including younger than 40 years of age, without clinical symptoms of dementia might show already AD-related changes in the brain even to a moderate degree (e.g., Braak stage III-IV of neurofibrillary degeneration[21] or early phases of A-beta deposition[22]). Neurofibrillary tau pathology can appear without A-beta deposition called primary age-related tauopathy (PART). The present study is the first to attempt a genome-wide analysis aiming to identify variants that are associated to healthy brain aging defined on both clinical and a neuropathological level. In an effort to characterise factors protecting against pathological age-related neurodegeneration, we implemented an approach that focuses on the extreme ends of the phenotypic distribution under study (i.e. a comparison of normal aging individuals >80 years versus individuals with pure AD, lacking other proteinopathies or vascular lesions, of the same age). In a community-based study involving demented and non-demented individuals each represented approximately 6-10% of the cohort[19], understanding such protective genetic factors may influence the adoption of strategies geared to prevent or decrease risk events that accelerate neurodegenerative changes in the susceptible aging brain when the processes are still reversible.

RESULTS

Following our alignment and quality controls steps, deep whole genome sequencing resulted in 10,908,452 variants in total, including 40,435 nonsynonymous, 475 stop-gain and 39 stop-loss variants annotated by RefSeq database (Table 1). Our analysis of variants previously implicated in AD replicated the positive association to ApoE allele (p-value 4.37e-05 (33% in Cases, 1.8% in Controls) and we were further able to pick up a neuroprotective effect for the ε2 allele (p-value of 0.02 - 0%Cases,17.9% Controls) (See Supplementary results). Looking into variants that have been previously implicated in longevity, we also picked up the association to *rs4420638* marker near *APOC1* gene which was previously implicated in a case-control GWAS conducted on 763 German centenarians and nonagenarians, and 1,085 controls (mean age 60 years)[16] (P value =2.00e-04, see Supplementary results for details).

To examine variants most likely to have a neuroprotective effect, we filtered for unique

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variants in at least 10 of the 28 neuropathologically "normal" controls, in our study but absent in all 12 cases. This resulted in 6,984 variants (Table S1), including 55 non-synonymous and one stopgain variant. Further removal of sequence artefacts and pseudogenes resulted in 14 prioritized exonic variants (Table 1, S2). A summary of these genes and their functions are provided in Table 2. Besides gene-coding variants, we also explored the potential involvement of variants in exons of non-coding RNAs. Again, as a filtering step, we focused on variants that are common in at least 10 "healthy" aged controls and absent from all cases. We identified eight such prioritized variants as shown in Table 3.

We proceeded to perform Sanger sequencing in order to confirm the whole genome sequencing results. Sanger sequencing was not able to confirm all of the whole genome sequencing called variants. We were thus able to further narrow down our search for genes that carry non-synonymous variants in at least 10 controls and none of the cases to only two genes; The Tectonin beta-propeller repeat containing 2 gene (*TECPR2*), with allele G on *rs10149146* present in 15 controls and no AD cases and the Cyclin-dependent kinase 2 interacting protein (*CINP*) with allele T on *rs7011* present in 13 controls and absent from all cases. The Fisher's exact test revealed statistically significant association to healthy brain aging with SNP *rs10149146* of *TECPR2* (P=4.055e-4). Interestingly, the two aforementioned SNPs lie physically close (86.16 kb) and are in high linkage disequilibrium (r^2=0.73). The genes carrying these variants have been previously found to share a bidirectional promoter thus being subject to common gene expression regulation[23]. Our analysis showed that transcription factors that potentially co-regulate the two genes (*TECPR2* and *CINP*) include OCT1 (Octamer-Binding Transcription Factor 1 - POU2F1), NMYC (N-myc proto-oncogene protein), USF (upstream transcription factor 1), TAX/CREB (Tax/cAMP response element-binding protein), HEN1 (Helix-Loop-Helix Protein 1), YY1 (Yin And Yang 1), and RFX1 (Regulatory Factor X1).

DISCUSSION

In the past two decades, evidence has accumulated in favour of the existence of genetic factors associated with AD-related pathology, via either GWAS studies with single-marker associations alone or by combining GWAS data with neuroimaging data, leading to the identification of 20 AD susceptibility loci[4–8,24–26]. However, very little is known about the protective genetic influences against neurodegeneration and in favour of healthy brain aging. Genetic influences of cognitive abilities and aging in old age are prominent and

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approximately 50% of this variance may be attributed to multiple genetic variants with small effects[27]. Candidate gene studies have highlighted the potential roles of genes such as *BDNF*, *COMT* and *DTNBP1* in either cognitive ability and normal function (such as learning and memory) or cognitive decline; yet associations for other genes remain inconclusive[28].

In this study, we applied a genome-wide sequencing analysis aiming at the identification of variants and genes that may be important for healthy aging, in an effort to characterise factors for the prevention of pathological age-related neurodegeneration. Importantly, based on detailed neuropathological evaluations, we were able to define two extremes of the phenotypic distribution of brain aging; old age individuals with minimal signs of neurodegeneration versus pure AD. Demonstrating the power of our approach, despite the small sample size, we were able to pick up the neuroprotective effect of the *APOE2* variant (p=0.02). We identified two non-synonymous exonic variants in two genes that are present in 53.6% ad 46% respectively of 28 healthy aging controls and absent from all 12 pure AD cases that we studied. The associated genes *TECPR2* and *CINP* appear to be "partner" genes in terms of regulation[23] and are associated with autophagy[29] and cell-cycle control[30] respectively. Notably, the transcription factors we found associated with these two genes have also previously been implicated in neurodegeneration; a *POU2F1 (OCT1)* intronic variant has been associated with AD and the *POU2F1* gene was found to be down-regulated in the hippocampus of AD brains but not healthy controls[31], USF has been found to interact with the Alzheimer amyloid beta-protein precursor gene[32], The beta-amyloid peptide mediates synapse loss through the CREB signaling pathway, while YY1 is an activator of beta-site amyloid precursor protein-cleaving enzyme 1 (BACE1), a prerequisite for the generation of beta-amyloid peptides, the principle constituents of senile plaques in the brains of patients with Alzheimer's disease (AD)[33]..

A variant in *TECPR2* (namely *rs10149146*, c.A2047G, p.I683V) was present in the highest number of our control individuals (n=15/28). *TECPR2* has been reported to be a positive regulator of autophagy and involved in hereditary spastic paraparesis[29]. Autophagy dysfunction is implicated in multiple progressive neurodegenerative diseases, and has been reported to play a major role in the pathogenesis of several neurodegenerative and muscular diseases, such as Huntington's Disease, AD and Parkinson's Disease, spinocerebellar ataxias, and amyotrophic lateral sclerosis[34]. Mutations in *TECPR2* have been also recently found to cause a subtype of familial dysautonomia comparable to hereditary sensory autonomic

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

neuropathy with intellectual disability[35]. *TECPR2* is also a possible candidate gene for human neuroaxonal dystrophies (NAD), which may be associated with aging and supplementary to numerous metabolic-toxic conditions[36].

On the other hand, 13 controls were found to share a variant in *CINP* (rs7011, CINP:NM_001177611:exon5:c.G536A:p.R179H, CINP:NM_032630:exon5:c.G491A:p.R164H) which was however absent from all of our AD cases. The protein encoded by *CINP* is a member of the cyclin-dependent kinase (CDK) family of Ser/Thr protein kinases and has been identified as a genomic maintenance protein [30]. It is a cell-cycle checkpoint protein, interacting with ATR-interacting protein to regulate ATR-dependent signaling, resistance to replication stress, and G2 checkpoint integrity [30]. Cdk2 is also a key regulator of the senescence control function of Myc and a potential therapeutic target for treating tumors derived by Myc or Ras [37]. Deregulation of proto-oncogenes like MYC and RAS augment tumor development although normally they function to promote normal cell growth and may have an impact on tissue regeneration[37]. An increasing body of studies show that cell cycle disturbances may play an early role in AD pathogenesis[38–40]. CDKs[41] propel tau phosphorylation. *In vitro* assays have indicated that ectopically re-expressed CDKs in neurons in AD phosphorylate tau in a similar manner to that found in AD *in vivo*[41]. Increased N-MYC and C-MYC expression in reactive astrocytes may also play a role in reactive astrocytosis in human neurodegenerative disorders [42].

Despite the relatively low number of cases examined, our strict criteria and by focusing on extreme ends of the phenotypic distribution of neuropathology-related variants, we were able to identify an autophagy and cell-cycle related gene to be associated with healthy aging specifically lacking pathological protein depositions and vascular lesions, which is unusual in the aging brain[19,20]. The implicated genes warrant further investigation and highlight the key role of autophagy and cell-cycle control in protecting from neurodegeneration. This is the first study to specifically attempt to find genetic variants that promote healthy brain aging based on neuropathological evaluations of post-mortem samples. Our approach contrasts those based only on clinical criteria but also neuropathology-based studies, which do not apply a holistic approach by examining the whole spectrum of proteinopathies and other brain lesions. In an aging world and with no existing cure for dementias, the need to increase our understanding of their etiology is imperative, Elucidating the physiological pathways that drive the shift from healthy brain aging to neurodegeneration, will lead to the identification of

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

targets for improved therapies and prevention strategies for AD and related neurodegenerative disorders.

METHODS

Subjects

We selected cases from the longitudinal aging transdanubian VITA study[19], which included individuals after written consent of either the patient or the family or legal representative of the patient. The Ethical Committee of the Medical University of Vienna gave approval for the genomic analysis (Nr: 140/2011). All experiments were conducted according to the principles expressed in the Declaration of Helsinki, and in accordance with relevant Austrian guidelines and regulations, including the right to object from participating in scientific research. The manuscript does not contain information or images that could lead to an identification of the individual or which could violate any personal rights. Bioinformatic processing was performed anonymously. Twelve participants with pure end-stage AD with only beta amyloid plaques (Thal phase 5) and neurofibrillary tangles (NFTs; Braak stage VI) in the brain and no other comorbidities (Lewy body pathology, TDP-43 proteinopathy, other tauopathies, vascular diseases etc.) participated (for details on neuropathological characterisation see[19]). A typically aging group of 28 participants aged 80 and above with minor amount of NFTs (Braak stage I but lack of A-beta plaques), not enough to be suggestive of AD and with no other neurodegenerative or vascular disorders were used as a control group.

Whole genome sequencing (WGS), processing of raw sequences and variant calling

DNA was extracted from fresh-frozen cerebellum and frontal cortex tissue. Whole genome sequencing was performed by the Biomedical Sequencing Facility at CeMM, using Illumina HiSeq 2000 machines with a read length was 2x100 base pairs in the paired-end configuration. The sequencing reads were aligned to the b37 reference genome with Epstein-Barr Virus sequence (BROAD 1000 Genomes GRCH37) using BWA-MEM aligner, and variant calling was done following the GATK[43] pipeline from GATK Best Practices available at the time of processing (http://www.broadinstitute.org/gatk/). Each sample was preprocessed to remove PCR and optical duplicates using Picard. In order to improve the SNP calling, INDEL realignment was done using IndelRealigner and recalibration of base quality scores was performed using baseRecalibrator from Genome Analyzer Toolkit, to eliminate systematic errors per lane. Once recalibrated, variant calling and recalibration was performed

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

on the merged samples using GATK version 3.1-1 Haplotype Caller. All samples achieved the goal of 30x genome-wide coverage with a mean coverage of 37.23x (4.47 Terabases of data). Only variants that passed the quality control filters (quality score QUAL>=30) and genotype threshold of 95% were kept for further analysis. Additional quality control steps were performed using vcftools (version 0.1.9)[44] and statistical analysis was done in R (version 3.0.1). *Variant Ranker* (http://paschou-lab.mbg.duth.gr/Software.html) was used to annotate and filter the variants. *Variant Ranker's* CaseControl filtering module utilizes SnpSift tool (CaseControl)[45] to count the number of homozygous, heterozygous and total variants in the case-control groups and implements ANNOVAR [46] for annotation of variants. Haploview[47] was used to perform Fischer's exact test.

Filtering of Variants and Case-Control analysis

RefSeq was used as the reference database for gene definitions. Our analysis primarily focused on functionally important exonic variants (i.e. non-synonymous/stopgain/stoploss/splicing effect) and variants present in exons of non-coding RNA genes (ncRNAs). In order to examine unique variants/mutations only present in the control samples with a likely neuroprotective effect, we filtered for variants that were carried by at least 10 control individuals (i.e. control subjects carried at least one alternative allele) and were absent in all 12 cases (i.e. cases carried only the reference alleles for the same locus in the human reference genome). Sequence artefacts and pseudogenes were removed from our analysis by examining gene exclusion lists in *Fuentes Fajardo 2012*[48]**.** Database for Annotation, Visualization and Integrated Discovery (DAVID)[49] was used for gene ontology analysis.

Sanger Sequencing

All controls carrying prioritized variants were sequenced by the Sanger custom sequencing method, using the BigDye Terminator v.3.1 chemistry on an ABI 3730XL genetic analyzer. Sequencing was performed on both DNA strands and sequences were both aligned and visually observed by at least two independent researchers.

ACKNOWLEDGEMENTS

This study was performed in the frame of the EU FP7 Project DEVELAGE (#278486; G.G.K.).

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author contributions. JA performed the statistical analysis, interpreted findings, and wrote the manuscript; PP and GGK were responsible for study design, supervised analysis and interpreted findings and wrote the manuscript; GGK supervised neuropathological evaluations; TS processed samples and performed sanger sequencing analysis; MG participated in the interpretation of results and wrote the manuscript; MC, TP, and CB performed whole genome sequencing analysis; SH, AR, PF, were responsible for the VITA study including clinical, pathological and neuropathological evaluations; all authors contributed to writing and revision of the manuscript

Competing financial interests Nothing to declare

REFERENCES

1. Aging, N. I. on. No Title. (2011). Available at: https://www.nia.nih.gov/research/publication/global-health-and-aging/preface.
2. Prince, M. *et al.* The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimers. Dement.* **9,** 63–75.e2 (2013).
3. Medway, C. & Morgan, K. Review: The genetics of Alzheimer's disease; putting flesh on the bones. *Neuropathol. Appl. Neurobiol.* **40,** 97–105 (2014).
4. 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).
5. 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).
6. Seshadri, S., AL, F., Ikram, M. & Al, E. Genome-wide analysis of genetic loci associated with alzheimer disease. *JAMA* **303,** 1832–1840 (2010).
7. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–1458 (2013).
8. Escott-Price, V. *et al.* Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. *PLoS One* **9,** e94661 (2014).
9. Mitsui, J. & Tsuji, S. Genomic aspects of sporadic neurodegenerative diseases. *Biochem. Biophys. Res. Commun.* **452,** 221–225 (2014).
10. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488,** 96–99 (2012).
11. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N.*

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Engl. J. Med. **368,** 107–116 (2013).

12. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* **368,** 117–127 (2013).
13. Herskind, A. M. *et al.* The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum. Genet.* **97,** 319–323 (1996).
14. Reed, T. & Dick, D. M. Heritability and validity of healthy physical aging (wellness) in elderly male twins. *Twin Res.* **6,** 227–234 (2003).
15. Ukraintseva, S. *et al.* Puzzling role of genetic risk factors in human longevity: 'risk alleles' as pro-longevity variants. *Biogerontology* **17,** 109–127 (2016).
16. Sebastiani, P. *et al.* Genetic signatures of exceptional longevity in humans. *PLoS One* **7,** (2012).
17. Lee, T., Henry, J. D., Trollor, J. N. & Sachdev, P. S. Genetic influences on cognitive functions in the elderly: a selective review of twin studies. *Brain Res. Rev.* **64,** 1–13 (2010).
18. Deelen, J. *et al.* Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23,** 4420–4432 (2014).
19. Kovacs, G. G. *et al.* Non-Alzheimer neurodegenerative pathologies and their combinations are more frequent than commonly believed in the elderly brain: a community-based autopsy series. *Acta Neuropathol.* **126,** 365–384 (2013).
20. Rahimi, J. & Kovacs, G. G. Prevalence of mixed pathologies in the aging brain. *Alzheimer's Research & Therapy* **6,** (2014).
21. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* **112,** 389–404 (2006).
22. Thal, D. R., Rub, U., Orantes, M. & Braak, H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology* **58,** 1791–1800 (2002).
23. Uchiumi, F., Larsen, S. & Tanuma, S. in *Advances in DNA Repair*
24. 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).
25. Sabuncu, M. R. *et al.* The association between a polygenic Alzheimer score and cortical thickness in clinically normal subjects. *Cereb. Cortex* **22,** 2653–2661 (2012).
26. Chauhan, G. *et al.* Association of Alzheimer's disease GWAS loci with MRI markers

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of brain aging. *Neurobiol. Aging* **36,** 1765.e7-16 (2015).

27. Harris, S. E. & Deary, I. J. The genetics of cognitive ability and cognitive ageing in healthy older people. *Trends Cogn. Sci.* **15,** 388–394 (2011).
28. Payton, A. The impact of genetic research on our understanding of normal cognitive ageing: 1995 to 2009. *Neuropsychol. Rev.* **19,** 451–477 (2009).
29. Oz-Levi, D. *et al.* Mutation in TECPR2 Reveals a Role for Autophagy in Hereditary Spastic Paraparesis. *American Journal of Human Genetics* **91,** 1065–1072 (2012).
30. Lovejoy, C. A. *et al.* Functional genomic screens identify CINP as a genome maintenance protein. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19304–9 (2009).
31. Taguchi, K. *et al.* Identification of hippocampus-related candidate genes for Alzheimer's disease. *Ann. Neurol.* **57,** 585–588 (2005).
32. Kovacs, D. M. *et al.* The upstream stimulatory factor functionally interacts with the Alzheimer amyloid beta-protein precursor gene. *Hum. Mol. Genet.* **4,** 1527–1533 (1995).
33. Nowak, K. *et al.* The transcription factor Yin Yang 1 is an activator of BACE1 expression. *J. Neurochem.* **96,** 1696–1707 (2006).
34. Oz-Levi, D., Gelman, A., Elazar, Z. & Lancet, D. TECPR2: A new autophagy link for neurodegeneration. *Autophagy* **9,** 801–802 (2013).
35. Heimer, G. *et al.* TECPR2 mutations cause a new subtype of familial dysautonomia like hereditary sensory autonomic neuropathy with intellectual disability. *Eur. J. Paediatr. Neurol.* **20,** 69–79 (2016).
36. Hahn, K. *et al.* TECPR2 Associated Neuroaxonal Dystrophy in Spanish Water Dogs. *PLoS One* **10,** e0141824 (2015).
37. Hydbring, P. & Larsson, L.-G. Cdk2: a key regulator of the senescence control function of Myc. *Aging (Albany. NY).* **2,** 244–250 (2010).
38. McShea, A., Harris, P. L., Webster, K. R., Wahl, A. F. & Smith, M. A. Abnormal expression of the cell cycle regulators P16 and CDK4 in Alzheimer's disease. *Am. J. Pathol.* **150,** 1933–1939 (1997).
39. Nagy, Z., Esiri, M. M., Cato, A. M. & Smith, A. D. Cell cycle markers in the hippocampus in Alzheimer's disease. *Acta Neuropathol.* **94,** 6–15 (1997).
40. Vincent, I., Rosado, M. & Davies, P. Mitotic mechanisms in Alzheimer's disease? *J. Cell Biol.* **132,** 413–425 (1996).
41. McShea, A. *et al.* Neuronal cell cycle re-entry mediates Alzheimer disease-type changes. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1772,** 467–472 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

42. Ferrer, I. & Blanco, R. N-myc and c-myc expression in Alzheimer disease, Huntington disease and Parkinson disease. *Brain Res. Mol. Brain Res.* **77,** 270–276 (2000).
43. Mckenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
44. Danecek, P. *et al.* The Variant Call Format and VCFtools. 3–5 (2011).
45. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* **6,** 80–92 (2012).
46. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
47. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–5 (2005).
48. Fuentes Fajardo, K. V *et al.* Detecting false-positive signals in exome sequencing. *Hum. Mutat.* **33,** 609–613 (2012).
49. Huang, D. W. *et al.* DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* **35,** (2007).
50. Lescai, F. *et al.* An APOE haplotype associated with decreased epsilon4 expression increases the risk of late onset Alzheimer's disease. *J. Alzheimers. Dis.* **24,** 235–245 (2011).
51. Jazwinski, S. M. *et al.* HRAS1 and LASS1 with APOE are associated with human longevity and healthy aging. *Aging Cell* **9,** 698–708 (2010).
52. Nonaka, M. *et al.* Region-specific activation of CRTC1-CREB signaling mediates long-term fear memory. *Neuron* **84,** 92–106 (2014).
53. Parra-Damas, A. *et al.* Crtc1 activates a transcriptional program deregulated at early Alzheimer's disease-related stages. *J. Neurosci.* **34,** 5776–5787 (2014).
54. Saura, C. A. & Valero, J. The role of CREB signaling in Alzheimer's disease and other cognitive disorders. *Rev. Neurosci.* **22,** 153–169 (2011).
55. Prawitt, D. *et al.* Identification and characterization of MTR1, a novel gene with homology to melastatin (MLSN1) and the trp gene family located in the BWS-WT2 critical region on chromosome 11p15.5 and showing allele-specific expression. *Hum. Mol. Genet.* **9,** 203–216 (2000).
56. Glendinning, J. I., Gillman, J., Zamer, H., Margolskee, R. F. & Sclafani, A. The role of T1r3 and Trpm5 in carbohydrate-induced obesity in mice. *Physiology & behavior* **107,**

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

50–58 (2012).

57. Li, L. *et al.* The gene for severe combined immunodeficiency disease in Athabascan-speaking Native Americans is located on chromosome 10p. *Am. J. Hum. Genet.* **62,** 136–144 (1998).
58. Ege, M. *et al.* Omenn syndrome due to ARTEMIS mutations. *Blood* **105,** 4179–4186 (2005).
59. Blanco, P., Sargent, C. A., Boucher, C. A., Mitchell, M. & Affara, N. A. Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain. *Mamm. Genome* **11,** 906–914 (2000).
60. Carrasquillo, M. M. *et al.* Genetic variation in PCDH11X is associated with susceptibility to late onset Alzheimer's disease. *Nature genetics* **41,** 192–198 (2009).
61. Gomes, A. V. Genetics of proteasome diseases. *Scientifica (Cairo).* **2013,** 637629 (2013).
62. Leevers, S. J., Vanhaesebroeck, B. & Waterfield, M. D. Signalling through phosphoinositide 3-kinases: the lipids take centre stage. *Curr. Opin. Cell Biol.* **11,** 219–225 (1999).
63. Chalbot, S. *et al.* Blood-cerebrospinal fluid barrier permeability in Alzheimer's disease. *J. Alzheimers. Dis.* **25,** 505–15 (2011).
64. Monaco, E. A. 3rd & Vallano, M. Lou. Role of protein kinases in neurodegenerative disease: cyclin-dependent kinases in Alzheimer's disease. *Front. Biosci.* **10,** 143–159 (2005).
65. Dinour, D. *et al.* Homozygous SLC2A9 Mutations Cause Severe Renal Hypouricemia. *Journal of the American Society of Nephrology : JASN* **21,** 64–72 (2010).
66. Hollingworth, P. *et al.* Genome-wide association study of Alzheimer's disease with psychotic symptoms. *Mol. Psychiatry* **17,** 1316–1327 (2012).
67. Yonekawa, S. *et al.* Sec16B is involved in the endoplasmic reticulum export of the peroxisomal membrane biogenesis factor peroxin 16 (Pex16) in mammalian cells. *Proc. Natl. Acad. Sci.* **108,** 12746–12751 (2011).
68. Hotta, K. *et al.* Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population. *J. Hum. Genet.* **54,** 727–31 (2009).
69. Ramanan, V. K. *et al.* FASTKD2 is associated with memory and hippocampal structure in older adults. *Mol. Psychiatry* 1–8 (2014). doi:10.1038/mp.2014.142

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Table 1. Summary statistics on whole genome sequencing data from 40 studied individuals after applying quality control steps.

CLASS OF VARIANTS	Number
downstream	66523
exonic	78443
intergenic	6412774
intronic	3845897
ncRNA_exonic	23776
ncRNA_intronic	314494
ncRNA_splicing	132
ncRNA_UTR3	1545
ncRNA_UTR5	324
splicing	327
upstream	65737
upstream;downstream	2369
UTR3	78640
UTR5	17422
UTR5;UTR3	35
exonic;splicing	14
TOTAL VARIANTS	**10908452**
EXONIC	
nonsynonymous	40435
stopgain	475
stoploss	39
synonymous	36526
unknown	982
TOTAL EXONIC	**78457**

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ANNOVAR definitions

Exonic - variant overlaps a coding region; **Splicing** - variant is within 2-bp of a splicing junction; **ncRNA** - variant overlaps a transcript without coding annotation in the gene definition;**UTR5** - variant overlaps a 5' untranslated region;**UTR3**- variant overlaps a 3' untranslated region; **Intron** - variant overlaps an intron;**Upstream**- variant overlaps 1-kb region upstream of transcription start site;**Downstream** - variant overlaps 1-kb region downtream of transcription end site;**Intergenic** - variant is in intergenic region

Table 2. Unique exonic variants (NS-nonsynonymous, stopgain) residing in protein-coding genes and present in at least 10 controls and absent from all 12 AD cases following whole genome-sequencing (WGS) analysis. Sanger sequencing validations are also shown..

Gene (protein name)	**Variant (rs number) and functional effect**	**No of controls (WGS)**	**No of controls (Sanger sequencing validation)**	**Protein Function**	**Evidence for association with neurodegenerative or other disorder**
APOE **(**apolipoprotein E)	rs7412 (NS)	10	4	ApOE encoded protein is a major apoprotein of the chylomicron. It binds to a specific peripheral cell and liver receptor, and is critical for the normal catabolism of triglyceride-rich lipoprotein constituents.	Late onset AD[50] and longevity[51]

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CRTC1 **(**CREB-regulated transcription coactivator 1)	rs3746266 (NS)	11	8	In the hippocampus, involved in late-phase long-term potentiation (L-LTP) maintenance at the Schaffer collateral-CA1 synapses. May be essential for dendritic growth of developing cortical neurons (By similarity). Regulating neuronal gene expression during learning and memory[52]	AD[53,54] "Role of Crtc1-dependent transcription on spatial memory formation and provide the first evidence that targeting brain transcriptome reverses memory loss in AD."
TRPM5 **(**transient receptor potential cation channel, subfamily M, member 5)	rs34350821 (NS)	11	7	Role in taste transduction	Beckwith-Wiedemann Syndrome[55] Obesity (in mice)[56]
DCLRE1C **(**DNA cross-link repair 1C)	rs12768894 (NS)	12	6	Nuclear protein involved in V(D)J recombination and DNA repair	Mutations in this gene can cause Athabascan-type severe combined immunodeficiency (SCIDA)[57] and Omenn Syndrome[58].
PCDH1	rs382235	10	5	Involved in neural	PCDH11X

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(protocadherin 1)	7 (NS)			cell adhesion, suggesting a possible role in neuronal development [59]	susceptibility in AD[60]
PSMF1 **(**proteasome (prosome, macropain) inhibitor subunit 1 (PI31)	rs2235587 (NS)	10	5	Control of proteasome function. Inhibits the hydrolysis of protein and peptide substrates by the 20S proteasome. Also inhibits the activation of the proteasome by the proteasome regulatory proteins PA700 and PA28.	Proteasome dysfunction is implicated in neurodegenerative disorders, muscular dystrophies cardiomyopathies, immune defects, metabolic diseases, and cancer[61]
PIK3AP1 **(**phosphoinositide-3-kinase adaptor protein 1)	rs12784975 (NS)	11	7	Responsible for the production of PI(3,4,5)P3. Contributes to the regulation of cell proliferation, metabolism, survival, cytoskeletal reorganization, and membrane trafficking [62]	
PLA2G2C (phospholipase A2, group IIC)	rs12139100 (stopgain)	12	8	Phospholipases are ubiquitously expressed and have diverse biological functions including roles in	Increased sPLA2 activity observed in AD[63]

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

				inflammation, cell growth, signaling and death and maintenance of membrane phospholipids.	
MEGF6 **(**multiple EGF-like-domains 6)	rs7513275 (NS)	13	6		
CINP **(**cyclin-dependent kinase 2 interacting protein)	rs7011 (NS)	17	13	Component of the DNA replication complex, as well as a genome-maintenance protein.	Dysthymic disorder. Cyclin-dependent kinase (Cdks) pathways may have a role in the neuronal loss that is responsible for AD[64]
SLC2A9 **(**solute carrier family 2 (facilitated glucose transporter), member 9)	rs6820230 (NS)	12	7	Members of this family play a significant role in maintaining glucose homeostasis. The encoded protein may play a role in the development and survival of chondrocytes in cartilage matrices.	Renal hypouricemia[65] Variants upstream of *SLC2A9* have been found to increase the risk in AD[66]
TECPR2 **(**tectonin beta-propeller	rs10149146 (NS)	19	15	Potential role as positive regulator of autophagy.	Spastic paraplegia[29] Autophagy impairment has been implicated in the

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

repeat containing 2)					pathogenesis of several neurodegenerative and muscle diseases, such as Huntington's, AD and Parkinson's diseases, spinocerebellar ataxias (SCAs), and amyotrophic lateral sclerosis[34]
SEC16B **(**SEC16 homolog B (*S. cerevisiae*))	rs3813649 (NS) and rs61745560 (NS)	13	7	Required for organization of transitional endoplasmic reticulum sites and protein export[67]	Obesity[68]

* Relevant citation provided if RefSeq summary is not available

** Reported associations to AD, other neurodegenerative disorders or aging are highlighted

bioRxiv preprint doi: https://doi.org/10.1101/157693; this version posted June 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Unique variants identified in exons of ncRNA genes in at least 10 controls and absent from all 12 AD cases following whole genome sequencing (WGS) analysis. Sanger sequencing validations are also shown.

Gene (associated variant rs number)	**No of controls (WGS)**	**N of controls (Sanger sequencing validation)**	**Function**	**Disease Associated**
LINC00485 (Long Intergenic Non-Protein Coding RNA 485) (rs11111387)	11	6		
C1orf140 (rs72746037)	11	4	uncharacterized	
DGCR5 (DiGeorge Syndrome Critical Region Gene 5) (rs9604914)	11	6		DiGeorge syndrome
LOC116437 (Long intergenic non-protein coding RNA 1257) (rs11061382, **rs58011062**)	10	6		
LINC00689 (rs12113517)	11	7		
LOC285768 (rs9378593)	10	3		
MIR3130-1 (rs2241347)	10	6		*FASTKD2* is associated with memory and hippocampal structure in older adults[69]