medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Large meta-analysis of genome-wide association studies expands knowledge of the genetic etiology of Alzheimer's disease and highlights potential translational opportunities

Céline Bellenguez[1,*,#], Fahri Küçükali[2,3,4*], Iris Jansen[5,6*], Victor Andrade[7,8*], Sonia Morenau-Grau[9,10,*], Najaf Amin[11,12], Benjamin Grenier-Boley[1], Anne Boland[13], Luca Kleineidam[7,8], Peter Holmans[14], Pablo Garcia[9,10], Rafael Campos Martin[7], Adam Naj[15,16], Yang Qiong[17], Joshua C. Bis[18], Vincent Damotte[1], Sven Van der Lee[5,6,19], Marcos Costa[1], Julien Chapuis[1], Vilmentas Giedraitis[20], María Jesús Bullido[10,21], Adolfo López de Munáin[10,22], Jordi Pérez-Tur[10,23], Pascual Sánchez-Juan[10,24], Raquel Sánchez-Valle[25], Victoria Álvarez[26], Pau Pastor[27], Miguel Medina[10,28], Jasper Van Dongen[2,3,4], Christine Van Broeckhoven[2,3,4], Rik Vandenberghe[29,30], Sebastiaan Engelborghs[31,32], Gael Nicolas[33], Florence Pasquier[34], Olivier Hanon[35], Carole Dufouil[36], Claudine Berr[37], Stéphanie Debette[36], Jean-François Dartigues[36], Gianfranco Spalletta[38], Benedetta Nacmias[39,40], Vincenzo Solfrezzi[41], Barbara Borroni[42], Lucio Tremolizzo[43], Davide Seripa[44], Paolo Caffarra[45], Antonio Daniele[46,47], Daniela Galimberti[48,49], Innocenzo Rainero[50], Luisa Benussi[51], Alesio Squassina[52], Patrizia Mecoci[53], Lucilla Parnetti[54], Carlo Masullo[55], Beatrice Arosio[56], John Hardy[57], Simon Mead[58], Kevin Morgan[59], Clive Holmes[60], Patrick Kehoe[61], Bob Woods[62], EADB, Charge, ADGC, Jin Sha[15,16], Yi Zhao[15,63], Chien-Yueh Lee[15,63], Pavel P. Kuksa[15,63], Kara L. Hamilton-Nelson[64], Brian W. Kunkle[64,65], William S. Bush[66], Eden R. Martin[64,65], Li-San Wang[15,63], Richard Mayeux[67,68], Lindsay A. Farrer[69], Jonathan L. Haines[66], Margaret A. Pericak-Vance[64,65], Ruiqi Wang[17], Claudia Satizabal[17], Bruce Psaty[18], Oscar Lopez[18], Florentino Sanchez-Garcia[70], Børge G. Nordestgaard[71,72], Anne Tybjærg-Hansen[72,73], Jesper Qvist Thomassen[74], Caroline Graff[75,76], Goran Papenberg[77], Hilkka Soininen[78,79], Miia Kivipelto[80,81], Annakaisa Haapasalo[82], Tiia Ngandu[80,83], Anne Koivisto[84,85], teemu kuulasmaa[86], Laura Molina Porcel[87,88], Johannes Kornhuber[89], Oliver Peters[90], Anja schneider[91], Nikolaos Scarmeas[92,93], Martin Dichgans[94], Lutz Froelich[95], Dan Rujescu[96], Janine Diehl-Schmid[97], Timo Grimmer[97], Matthias Schmid[98], Markus M Möthen[99], Edna Grünblatt[100,101], Julius Popp[102], Norbert Scherbaum[103], Shima Mehrabian[104], Jürgen Deckert[105], Dag Aarsland[106], Geir Selbæk[107], Ingvild Saltvedt[108], Srdjan Djurovic[109], Henne Holstege[5], Yolande A.L. Pijnenburg[5], John Van Swieten[12], Inez Ramakers[110], Aad Van der Lugt[111], Jurgen A.H.R Claassen[111], Geert Jan Biessels[111], Philip Scheltens[5], Carmen Antúnez[1112], Pablo Mir[10,113], Luis Miguel Real[114], Jose María García-Alberca[115], Gerard Piñol-Ripoll[116,117], Guillermo Garcia-Ribas[118], Manuel Serrano-Ríos[119], Julie Williams[14,120], Perminder Sachdev[121], Philippe Amouyel[1], Karen Mather[121], Frank Jessen[122,123], Alexandre de Mendonça[124], Jakub Hort[125,126], Magda Tsolaki[127], Ruth Frikke-Schmidt[72,73], Jordi Clarimon[10,28], Jean-François Deleuze[13], Sudha Seshadri[128,129], Gerald Schellenberg[15,63], Giacomina Rossi[130], Ole Andreassen[131,132], Martin Ingelsson[20], Mikko Hiltunen[86], Kristel Sleegers[2,3,4], Cornelia Van Duijn[11,12,**], Rebecca Sims[14,120**], Wiesje M. Van der Flier[5,**], Agustin Ruiz[9,**], Alfredo Ramirez[7,**], Jean-Charles Lambert[1,**,#]

* co-first
** co-last
\# corresponding author
Céline Béllenguez: celine.bellenguez@pasteur-lille.fr
Jean-Charles Lambert: jean-charles.lambert@pasteur-lille.fr

1. Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167-RID-AGE-Facteurs de risque Et déterminants moléculaires des maladies liées au vieillissement, Lille, France.
2. VIB Center for Molecular Neurology, VIB, Antwerp, Belgium.
3. Laboratory of Neurogenetics, Institute Born - Bunge, Antwerp, Belgium
4. Department of Biomedical Sciences, University of Antwerp, B, Antwerp, Belgium
5. Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

6. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University, Amsterdam, the Netherlands
7. Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, 50937 Cologne, Germany.
8. Department of Neurodegeneration and Geriatric Psychiatry, University of Bonn, 53127 Bonn, Germany.
9. Research Center and Memory clinic Fundació ACE, Institut Català de Neurociències Aplicades, Universitat Internacional de Catalunya, Barcelona, Spain
10. CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain
11. Nuffield Department of Population Health, University of Oxford, UK
12. Department of Epidemiology, Erasmus Medical Center Rotterdam, the Netherlands.
13. Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France.
14. Division of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK
15. Department of Biostatistics, Epidemiology, and Informatics; University of Pennsylvania Perelman School of Medicine, USA
16. Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, USA
17. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.
18. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA.
19. Department of Clinical Genetics, VU University Medical Centre, Amsterdam, The Netherlands
20. Department of Public Health and Caring Sciences/Geriatrics, Uppsala University, Uppsala, Sweden
21. Centro de Biología Molecular Severo Ochoa (CSIC-UAM (Madrid), Spain
22. Institute Biodonostia, University of Basque Contry (EHU-UPV, San Sebastián), Spain
23. Unitat Mixta de Genètica i Neurologia, Institut d'Investigació Sanitària La Fe (València), Spain.
24. Neurology Service, University Hospital Marqués de Valdecilla (Santander)-IDIVAL-UC), Spain
25. Alzheimer's disease and other cognitive disorders, Neurology Department, at Hospital Clínic, IDIBAPS (Barcelona), Spain
26. Molecular Genetics Laboratory, at the Hospital Universitario Central de Asturias (Oviedo), Spain
27. Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mútua de Terrassa (Barcelona), Spain
28. Fundación CIEN (Madrid), Spain
29. Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Belgium
30. Neurology Department, University Hospitals Leuven, Leuven, Belgium
31. Department of Neurology and Center for Neurosciences, UZ Brussel and Vrije Universiteit Brussel (VUB), Brussels, Belgium
32. Reference Center for Biological Markers of Dementia (BIODEM), Institute Born-Bunge, University of Antwerp, Antwerp, Belgium
33. Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics and CNR-MAJ, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France
34. Univ Lille, Inserm U1172, CHU, DISTAlz, LiCEND, F-59000 Lille, France
35. APHP, Hôpital Broca, Paris, France
36. Inserm, Bordeaux Population Health Research Center, UMR 1219, Univ. Bordeaux, ISPED, CIC 1401-EC, Univ Bordeaux, F-33000 Bordeaux, France
37. Univ. Montpellier, Inserm U1061, Neuropsychiatry: epidemiological and clinical research, PSNREC, 39 avenue Charles Flahault, BP 34493, 34093, Montpellier cedex 05, France.
38. Laboratory of Neuropsychiatry, IRCCS Santa Lucia Foundation, Rome, Italy
39. Department of Neuroscience, Psychology, Drug Research and Child Health University of Florence, Florence Italy
40. IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy
41. "Cesare Frugoni" Internal and Geriatric Medicine and Memory Unit, University of Bari "Aldo Moro" , Bari, Italy.
42. Centre for Neurodegenerative Disorders, Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

43. School of Medicine and Surgery, University of Milano-Bicocca and Milan Center for Neuroscience, Italy
44. Complex Structure of Geriatrics, Department of Medical Sciences, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG), Italy.
45. Unit of Neuroscience, DIMEC, University of Parma, Parma, Italy
46. Institute of Neurology, Catholic University of Sacred Heart, Rome, Italy
47. Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy
48. University of Milan, Dino Ferrari center, Milan, Italy
49. Fondazione IRCCS Ca' Granda, Ospedale Policlinico, Milan, Italy
50. Department of Neuroscience "Rita Levi Montalcini", University of Torino, Torino, Italy
51. Molecular Markers Laboratory, IRCCS Instituto Centro San Giovanni di Dio Fatebenefratelli, Brescia
52. Department of Biomedical Sciences, Section of Neuroscience and Clinical Pharmacology, University of Cagliari, Italy
53. Department of Medicine, Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy.
54. Centre for Memory Disturbances, Lab of Clinical Neurochemistry, Section of Neurology, University of Perugia, Italy
55. Institute of Neurology, Catholic University of the Sacred Heart, School of Medicine
56. Department of Clinical Sciences and Community Health, University of Milan, Italy
57. Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
58. MRC Prion Unit at UCL, UCL Institute of Prion Diseases, London, UK
59. Human Genetics, School of Life Sciences, University of Nottingham, UK
60. Clinical and Experimental Science, Faculty of Medicine, University of Southampton, Southampton, UK
61. Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS16 1LE, UK
62. School of Health Sciences, Bangor University, UK
63. Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
64. John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA
65. Dr. John T. Macdonald Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA
66. Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA
67. Taub Institute for Research on Alzheimer Disease and the Aging Brain, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA;
68. Gertrude H. Sergievsky Center, Vagelos College of Physicians and Surgeons, Columbia University, New York, USA
69. Departments of Medicine, Neurology, Epidemiology, and Biostatistics, Boston University School of Medicine, Boston, MA, USA
70. Servicio de Inmunología. Hospital Universitario de Gran Canaria Dr.Negrín, Spain
71. Department of Clinical Biochemistry, Herlev-Gentofte Hospital, Herlev, Denmark.
72. Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark
73. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
74. Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark.
75. Karolinska Institutet, Center for Alzheimer Research, Department NVS, Division of Neurogeriatrics, 171 64 Stockholm, Sweden
76. Unit for Hereditary Dementias, Theme Aging, Karolinska University Hospital-Solna, 171 64 Stockholm Sweden
77. Aging Research Center, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, Sweden
78. Institute of Clinical Medicine, neurology, University of Eastern Finland, Kuopio, Finland
79. Neurocenter, Neurology, Kuopio University Hospital, Kuopio, Finland
80. Division of Clinical Geriatrics, Center for Alzheimer Research, Care Sciences and Society (NVS), Karolinska Institutet, Stockholm, Sweden
81. Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland
82. A.I Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

83. Public Health Promotion Unit, Finnish Institute for Health and Welfare, Helsinki, Finland
84. Department of Neurology, University of Eastern Finland, Kuopio, Finland.
85. Department of Neurosciences, University of Helsinki, Helsinki, Finland
86. Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland
87. Neurological Tissue Bank - Biobanc- Hospital Clinic -IDIBAPS, Barcelona, Spain
88. Alzheimer's disease and other cognitive disorders Unit. Neurology Department, Hospital Clinic , Barcelona, Spain
89. Department of Psychiatry and Psychotherapy, Universitätsklinikum Erlangen, and Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany.
90. Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Psychiatry and Psychotherapy, Hindenburgdamm 30, 12203 Berlin, Germany
91. Department for Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany
92. Taub Institute for Research in Alzheimer's Disease and the Aging Brain, The Gertrude H. Sergievsky Center, Department of Neurology, Columbia University, New York, NY
93. 1st Department of Neurology, Aiginition Hospital, National and Kapodistrian University of Athens, Greece
94. German Center for Neurodegenerative Diseases (DZNE), Munich, Germany
95. Department of Geriatric Psychiatry Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, 68159, Mannheim, Germany
96. University Clinic and Outpatient Clinic for Psychiatry, Psychotherapy and Psychosomatics, Martin Luther University Halle-Wittenberg, Halle, Germany.
97. Department of Psychiatry and Psychotherapy, Technical University of Munich, Germany.
98. German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany.
99. Venusber-Campus 1, 53127 Bonn
100. Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, University of Zurich, Zurich, Switzerland
101. Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland
102. Institute for Regenerative Medicine, University of Zürich, Switzerland
103. Virchowstr. 174, 45147 Essen, Germany
104. 1, Georgi Sofiiski, str; 1431 Sofia, Bulgaria
105. Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital Würzburg, Margarete-Höppel-Platz 1, 97080 Würzburg, Germany
106. Department of Old Age Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK
107. Norwegian National Advisory Unit on Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway
108. Department of geriatrics, Clinic of medicine, St Olavs hospital, University hospital of Trondheim, Norway
109. Department of Medical Genetics, Oslo University Hospital, Oslo, Norway and NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway
110. Maastricht University, Department of Psychiatry & Neuropsychologie, Alzheimer Center Limburg, Maastricht, the Netherlands
111. Radboudumc Alzheimer Center, Department of Geriatrics, Radboud University Medical Center, Nijmegen, the Netherlands
112. Unidad de Demencias, Hospital Clínico Universitario Virgen de la Arrixaca, Murcia, Spain.
113. Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología. Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad dSevilla, Seville, Spain.
114. Unidad Clínica de Enfermedades Infecciosas y Microbiología. Hospital Universitario de Valme, Sevilla, Spain.
115. Alzheimer Research Center & Memory Clinic, Andalusian Institute for Neuroscience, Málaga, Spain.
116. Unitat Trastorns Cognitius, Hospital Universitari Santa Maria de Lleida, Lleida, Spain.
117. Institut de Recerca Biomédica de Lleida (IRBLLeida), Lleida, Spain.
118. Hospital Universitario Ramon y Cajal, IRYCIS, Madrid.
119. Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas, CIBERDEM, Spain, Hospital Clínico San Carlos, Madrid, Spain.
120. UK Dementia Research Institute at Cardiff, Cardiff University, Cardiff, UK.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

121. Psychiatry Postgraduate Co-Convenor, CHeBA (Centre for Healthy Brain Ageing), UNSW Medicine, School of Psychiatry, Sydney, Australia
122. Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, 50937 Cologne, Germany
123. German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany
124. Faculty of Medicine, University of Lisbon, Portugal.
125. Memory Clinic, Department of Neurology, Charles University, 2nd Faculty of Medicine and Motol University Hospital, Czech Republic
126. International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic
127. 1st Department of Neurology, Medical School, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece.
128. Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA
129. Framingham Heart Study, Framingham, MA, USA.
130. Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy.
131. Norwegian Centre for Mental Disorders Research (NORMENT), Institute of Clinical Medicine, University of Oslo, Oslo, Norway
132. Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Deciphering the genetic landscape of Alzheimer's disease (AD) is essential to define the pathophysiological pathways involved and to successfully translate genomics to potential tailored medical care. To generate the most complete knowledge of the AD genetics, we developed through the European Alzheimer's Disease BioBank (EADB) consortium a discovery meta-analysis of genome-wide association studies (GWAS) based on a new large case-control study and previous GWAS (in total 39,106 clinically diagnosed cases, 46,828 proxy-AD cases and 401,577 controls) with the most promising signals followed-up in independent samples (18,063 cases and 23,207 controls). In addition to 34 known AD loci, we report here the genome-wide significant association of 31 new loci with the risk of AD. Pathway-enrichment analyses strongly indicated the involvement of gene sets related to amyloid and Tau, but also highlighted microglia, in which increased gene expression corresponds to more significant AD risk. In addition, we successfully prioritized candidate genes in the majority of our new loci, with nine being primarily expressed in microglia. Finally, we observed that a polygenic risk score generated from this new genetic landscape was strongly associated with the risk of progression from mild cognitive impairment (MCI) to dementia (4,609 MCI cases of whom 1,532 converted to dementia), independently of age and the *APOE* ε4 allele.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

The objectives of translational genomics and subsequent personalized medicine are to define prognosis/diagnosis markers of a disease and to adapt treatments at the individual level. Such approaches have already been successful in cancers and promising approaches are emerging for tailored treatments for diabetes[1]. However, in many other pathologies such as dementia, personalized approaches are still a concept and have proven difficult to develop.

Indeed, dementia is a term that encompasses an array of complex phenotypes in which the main symptom is the progressive decline of cognitive performance as observed in Alzheimer's disease (AD), the most common type. The dementia stage in AD is the culmination of a long, progressive, and silent process that is followed by intermediate pathological changes leading ultimately to cognitive decline and dementia. Importantly, the concept of AD has changed to recognize that AD is a continuum with a long preclinical phase of Subjective Cognitive Decline (SCD), a stage of mild cognitive impairment (MCI), and a dementia phase. The preclinical phase may offer unique opportunities for prevention of AD through early detection of AD pathology and application of pharmacological treatment with disease-modifying drugs that are still under development.

In this context, translational genomics may be of particular interest in AD since this disease exhibits a particularly high heritability, estimated between 60 and 80%[2]. Indeed, since the first large genome-wide association studies (GWAS) published in 2009[3,4], many loci/genes have been associated with the risk of developing AD[5]. These genetic findings have contributed to the identification of pathways and networks underlying AD, specifically implicating immunity, cholesterol processing, endocytosis, and more recently, the role of Aβ and Tau in the pathogenesis of common forms of AD[6]. Indeed, it is expected that the discovery of genetic risk factors for AD will reveal additional relevant pathogenic pathways operating in AD. Mounting evidence now suggests that AD is a disease in which multiple components combine to trigger the disease, beyond the dominant “amyloid cascade hypothesis”. This observation may also indicate that preponderant deleterious pathway(s) might be differentially involved at the individual level. For instance, therapies targeting the APP metabolism pathway may not be effective if this pathway is relatively unimportant for an individual with a particular genetic profile. If true, this would imply the importance of pursuing multiple therapies that target different genetically driven-pathways. Once a range of treatment options become available, an individualized model of AD pathology would be feasible, such that polytherapies and personalised medicine approaches can be developed and applied.

In a personalized medicine framework, it will be essential to translate large-scale genomic information into useful tools for personalized risk prediction and subsequent potential tailored intervention, for example through polygenic risk scores (PRS). Generating PRSs has been regarded as a reasonable solution to summarize genome-wide genotype data into a single variable that measures the genetic contribution to a trait or a disease for a particular individual. Herein, genomic information offers a unique opportunity for early detection. However, a large part of the AD genetic component is still unknown and several loci/genes also need to be confirmed as genuine genetic risk factors.

As a consequence, strong efforts are still needed to characterize the genetic architecture of AD, with the objectives to identify critical pathways and construct powerful PRSs for the disease. Within this background, increasing the size of GWAS data is an obvious way to facilitate the characterization of new genetic risk factors as observed in many other multifactorial diseases. In addition, since rare variants might explain a large proportion of the missing heritability, improving their analyses is also mandatory. Taking into account these two major points, we developed the European Alzheimer’s Disease BioBank (EADB) consortium grouping together the main European GWAS consortia already working on AD and a new dataset of 20,464 AD cases and 22,244 controls collated from 15 European

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

countries (Belgium, Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Greece, Italy, Portugal, Spain, Sweden, Switzerland, The Netherlands and The United Kingdom). We then benefited from the Trans-Omics for Precision Medicine (TOPMed) imputation panel based on whole-genome sequencing of 62,784 individuals in order to increase the number of variants tested and to improve the imputation quality of the rare variants[7]. The EADB results were meta-analysed with a proxy-AD GWAS performed in the UK Biobank. The best hits ($p \leq 10^{-5}$) generated from this meta-analysis step were then replicated in a large set of samples from the ADGC and CHARGE consortia.

In addition, in EADB, we collected an independent longitudinal cohort of 4,609 MCI cases of whom 1,532 converted to dementia. This provided us with the unique opportunity to test the association between a PRS we generated from our GWAS data and the risk of progression to dementia/AD with the objective to translate genomic information into personalized risk profiles for early detection of AD risk.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

RESULTS

GWAS analysis

The EADB Stage I (GWAS meta-analysis) was based on 39,106 clinically diagnosed cases, 46,828 proxy-AD cases and 401,577 controls (Supplementary Tables 1 and 2) and on 21,101,114 variants after quality control. Genomic inflation factors (λ) were slightly inflated (λ =1.08 overall and 1.17 when restricted to variants with minor allele frequency (MAF) above 1% (see Supplementary Figure 1 for a quantile–quantile (QQ) plot). However, linkage disequilibrium score (LDSC) regression estimate[8] indicated that the majority of this inflation was due to a polygenic signal, with the intercept being close to 1 (intercept=1.05, s.e=0.01 versus λ=1.2 on the variants considered in the LDSC analysis).

We selected all variants with *P* value less than 1×10^{-5} in Stage I. We defined non-overlapping regions around those variants, excluded the region corresponding to *APOE*, and sent the remaining variants for follow-up (Stage II, n=11,390, see Methods) in a large set of samples from the ADGC and CHARGE consortia (18,063 cases and 23,207 controls). A signal was considered as genome-wide significant if nominally replicated (*P* value≤0.05) in the same direction in the Stage I and Stage II analyses and if associated with AD risk with a *P* value less than 5×10^{-8} in the Stage I + Stage II meta-analysis. In addition, we applied a PLINK clumping procedure[9] to define potential independent hits from the Stage I results (see Methods). After validation by conditional analyses (see Supplementary Information and Supplementary Tables 3 and 4), this approach led us to confirm 38 signals in 34 loci already known to be associated with the risk of developing AD in the main previous AD GWASs[6,10–15] (Table 1) and to propose 31 new loci (Table 2 and Supplementary Figures 2-24). Five of these loci (APP, CCDC6, NCK2, TSPAN14 and Sharpin) were already reported in two preprints using GWAS data included in our study[16,17]. Besides, the NDUFAF6 and IGH loci were previously reported in a gene-wide analysis[18]. Of note, the magnitudes of associations in Stage I were highly similar to those observed if we restricted the Stage I to clinically-diagnosed cases, hereafter denoted as diagnosed cases only analysis (Supplementary Table 5 and Supplementary Figure 25). In addition, we did not detect any signal which may be mainly brought by the proxy-AD cases (Supplementary Table 5). Of note, we also provided a list of four loci with a genome-wide significant signal in the Stage I + Stage II analysis, but failing in Stage II (Supplementary Table 6).

Pathway analyses

To evaluate the biological significance of this new AD genetic landscape, we first performed pathway enrichment analyses. 150 gene sets were significant after multiple testing correction (q≤0.05, see Methods) in the Stage I (Supplementary Table 7), with the 20 most significant pathways shown in Table 3. The most significant gene sets relate to amyloid and Tau, with many of the other significant gene sets relating to lipids and immunity. Notably, there are gene sets related to macrophage activation and microglial cell activation. We also assessed whether enrichment pathway analyses were sensitive to the inclusion of the proxy-AD cases in Stage I. When analyses were performed limited to the diagnosed cases only, 69 gene sets were significant (q≤0.05) (Supplementary Table 7). Of these 69 gene sets, 53 reached q≤0.05 and all 69 reached p≤0.05 in the full Stage I that includes proxy-AD cases. This indicates that including proxy-AD cases did not mask disease-relevant biological information. We also repeated the enrichment analysis by using a window of 35kb upstream and 10kb downstream to assign variants to genes or by removing all 70 genes within 1Mb of APOE. Results were consistent between analyses (Supplementary Table 7).

We then performed a single-cell expression enrichment analysis in MAGMA using human data from the Allen Brain Atlas dataset (see Methods). Two complementary measures were used for each cell type: average gene expression per nucleus (Av. exp) and percentage of nuclei in a cell type expressing each gene (% Cell exp). Gabaergic/glutamergic neurons, astrocytes, oligodendrocytes, microglia and endothelial cells were analyzed and only

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

microglia expression reached significance after correcting for multiple testing in the two measurements (FDR≤0.05; Av. Exp.; $p=4.60x10^{-4}$ and % Cell. Exp. $p=5.59x10^{-8}$, Supplementary Table 8), with increased expression corresponding to more significant association with AD risk. A similar result was also observed using the mouse single cell dataset from Skene et al[19] (Supplementary Table 9).

We tested whether the relationship between microglia expression and association with AD risk was specific to particular areas of biology by using MAGMA. In particular, we tested for interaction between expression and pathway membership of each of the 91 significant pathways containing at least 25 genes with measured expression (Supplementary Table 10). Among several significant interactions, the most significant one was detected between GO:1902991 (regulation of amyloid precursor protein catabolic process) and gene expression level in microglia after multiple testing correction (q-values=$1.2x10^{-12}$ and $8.3x10^{-8}$ for % Cell Exp. and Av. Exp. respectively). This interaction is still significant when the APOE locus is removed (q-values= $3.8x10^{-11}$ and $7.8x10^{-4}$ for % Cell Exp and Av. Exp. respectively). This observation indicates that among the amyloid-beta gene sets showing an overall enrichment for AD association signal, it is the genes in these pathways with highest microglia expression that show the most association, suggesting a functional relationship between microglia and APP/Aβ pathways. A complete list of genes in APP/Aβ pathways GO:1902991 (but also GO:1902003), along with their GWAS significance and microglia expression, is given in Supplementary Table 11.

Gene prioritization

In order to prioritize candidate genes in the new loci, we considered the "nearest" gene from the lead variant and the genes exhibiting AD-related modulations within a region of 1 Mb around the lead variant according to those criteria: (i) expression and splicing quantitative trait loci (eQTLs and sQTLs) and colocalization analyses combined with transcriptome-wide association studies on expression and splicing (eTWAS and sTWAS) in AD-relevant brain regions; (ii) genetic-driven methylation as a biological mediator of genetic signals in blood (MetaMeth). We also considered those additional criteria: (i) the functional impact of gene under-expression on APP metabolism[20]; (ii) methylation QTL (mQTL) and histone acetylation QTL (haQTL) effects of the lead variants in dorsolateral prefrontal cortex (DLPFC)[21] and (iii) additional eQTL effects of these variants in naïve state monocytes and macrophages[22–27]. All the results were summarized in Fig. 2 and a full description of how the genes were prioritized is reported in the Supplementary Information (see also Supplementary Tables 12-21 and Supplementary Fig. 26-34).

Although the lead variants did not fall within a gene, for eight of the novel loci, brain molecular QTL, TWAS, blood MetaMeth and/or APP metabolism results exclusively supported the genes nearest to the lead variant: *OTULIN* (locus 4), *RASA1* (locus 5), ICA1 (locus 9), *TMEM106B* (locus 10), ABCA1 (locus 15), *CTSH* (locus 23), *MAF* (locus 25) and *SIGLEC11* (locus 28). Three other "nearest" genes in the new loci can be prioritized since the lead variant corresponds to a predicted deleterious missense variant within the gene itself: *MME* (locus 2), *FDFT1* (locus 12), and *SHARPIN* (locus 14). For *SHARPIN*, we found additional evidence that AD risk in the locus is associated with SHARPIN expression and splicing events (Supplementary Tables 17-18 and Supplementary Fig. 31-32). Finally, for six loci: *NCK2* (locus 1), *RASGEF1C* (locus 6), *HS3HT5* (locus 7), *UMAD1* (locus 8), *C1S* (locus 19) and *APP* (locus 31), none of the candidate genes could be prioritized based on genetic-driven expression, splicing or methylation analyses, therefore we considered that their proximity to the lead variant was in favor of their prioritization but at this stage at a lower level of confidence. Of note, *APP* is an obvious candidate gene, but *CYYR1-AS1* in between *APP* and *ADAMTS1* might also be of interest (Supplementary Tables 17, 21 and 31).

The remaining 14 novel loci present a more complicated pattern; several genes exhibit AD-related modulations in the same locus, and/or the prioritized gene is not the nearest protein coding gene. First, we could efficiently prioritize candidate risk genes in 4 additional loci:

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

EGFR (locus 11), *TSPAN14* (locus 17), *BLNK (locus 18)* and *GRN* (locus 27). For instance, in locus 11, EGFR is a likely candidate gene because its eQTL signals colocalize with the AD risk association signal, and its fine-mapped eTWAS hits (with FOCUS PIP values of ~1) associate predicted increased EGFR expression with increased AD risk (Fig 3; Supplementary Tables 15,17; supplementary Fig. 29, 31). In the complex locus 17, TSPAN14 was identified as the candidate risk gene as it exhibited numerous AD-related expression and splicing modulations, including novel cryptic complex splicing events that we identified and experimentally confirmed (Fig. 3; Supplementary Tables 13-19 and supplementary Fig. 26,27,29-32).

We did not clearly identify a single candidate in the remaining 10 loci. However, current evidence points towards the following candidate genes: i) *DGKQ, SCL26A1* and *IDUA* in the complex locus 3; ii) *CCDC6* in locus 16; iii) *SNX1* in locus 22 (importantly, we previously determined that *SNX1* and *APH1B* GWAS signals in this region were independent, see Supplementary Tables 3 and 4); iv) *INO80E*, *DOC2A*, *TBX6* and *YPEL3* in the most complex new locus 24 (*ALDOA* locus); v) *LIME1* and *RTEL* in locus 30. In locus 13, among the 4 genes, *NDUFAF6* exhibited the highest number of hits but *TP53INP1* was also of interest. In locus 21, a recent report pointing to an increased burden of rare variants in *ATP8B4* in AD patients prioritizes this gene[28]. For the locus 29, we consider *LILRB2* as a plausible risk gene according to bibliographical data[29,30]. Finally, we were not able to prioritize a gene in the complex IGH cluster (locus 20), nor in the *PRDM7* locus (locus 26).

Polygenic risk score

In order to explore the effect of a genetic burden on progression from mild cognitive impairment to AD-dementia, a PRS based on the genetic data generated above (see Methods and Supplementary Table 22) was constructed and tested in several longitudinal cohorts of MCI cases (Supplementary Table 23).
We observed a significant association of this PRS with the risk of progressing to any type of dementia (HR=1.028 per average risk variant, 95%CI [1.022-1.033], $p=4.93x10^{-7}$) or with the risk of progression to AD dementia only (HR=1.033 per average risk variants, 95%CI (1.027-1.039), $p=8.6x10^{-8}$) after adjustment on age, sex, principal components and the number of APOE-e4 alleles (Figure 4). Unadjusted analysis, analysis adjusted for age, sex and PCs only and coding non-AD converters as censored cases in the progression to AD dementia analysis did not change the results (see Supplementary table 24). Importantly, association of the PRS with progression risk does not seem to be modified by the presence of APOE-ε4 since we did not find any significant interaction between the number of APOE-e4 alleles and the PRS whatever the model tested.
Of note, the number of APOE-e4 alleles itself exerted a strong effect on the progression to all-cause of dementia (HR=1.64, 95%CI [1.51-1.78], $p=1.2x10^{-33}$) and AD-dementia (HR=1.79 [1.64-1.96], $p=3.6x10^{-38}$). This effect corresponds to carry 18 average risk variants coded in our PRS.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

DISCUSSION

This meta-analysis combining a new large case-control study and previous GWAS identified 69 independent genetic risk factors for AD, 38 previously reported in published GWAS and 31 corresponding to novel signals, strongly expanding our knowledge of the complex genetic landscape of Alzheimer's disease. We leveraged genetic, functional, and prior literature to nominate credible candidate genes at each new locus with a particular emphasis on transcriptional regulation, methylation, and APP metabolism datasets. A short description of these genes and their potential implication in AD are described in the Supplementary Information.
Remarkably, our meta-analysis and the characterization of these new loci clarify the global picture of AD etiology. For instance, pathway enrichment analyses remove ambiguities concerning the involvement of Tau binding proteins and APP/Aβ metabolism as major actors of the AD processes, beyond the levels of certainty previously described[6].

In addition, beyond the genetic risk factors already known to be involved in the APP metabolism, i.e. *ADAM10*, *APH1B* and *FERMT2*[20], we also proposed five candidate genes susceptible to modulate this metabolism (*DGKQ*, *RASA1*, *ICA1*, *DOC2A* and *LIME1)*. Although further investigations are needed to determine their exact implications, our data clearly support the central role of the APP/Aβ metabolism/functions in the pathophysiological process of late-onset forms of AD. Of note, none of these genes were included in the GO:1902992 pathway (negative regulation of amyloid precursor protein catabolic process) we characterized as the most enriched in potential AD genetic risk factors (Table 3).
These enrichment pathway analyses also confirmed the involvement of innate immunity and microglial activation in AD (Table 3). In addition, single-cell expression enrichment analysis also highlighted genes expressed in microglia (Supplementary Tables 8 and 9). Finally, 9 of our prioritized genes, i.e. *OTULIN*, *RASGEF1C*, *TSPAN14*, *BLNK*, *ATP8B4*, *MAF*, *GRN*, *SIGLE11C* and *LILRB2*, appeared to be mainly or almost only expressed in microglia (Fig. 2). However, at this time, only *GRN* is currently referenced in the microglial cell activation pathway (GO:0001774; Table 3). This suggests that enrichment of pathways involving microglia may be underestimated in our current analysis and further works will be clearly needed to determine whether and how these 9 genes may be involved in microglia function/activation. Several publications have already demonstrated involvement of the corresponding proteins in microglia function/activation (see Supplementary Information) and importantly three, i.e. *GRN*, *SIGLEC11* and *LILRB2* have also been linked to Aβ peptides/amyloid plaques[29,31,32]. Taking into account the already known genetic risk factors primarily expressed in microglia, i.e. *INPPD5*, *TREM2*, *SPI1*, *MS4A4A*, *SPPL2A*, *PLCγ2* and *ABI3* (Supplementary Figure 35), this means that at least 25% of the genetic risk loci described in this paper are credibly linked to AD-related microglia dysfunctions. Importantly, *TREM2*, *PLCg2*, *ABI3* and *INPP5D* were also characterized as microglia Aβ response proteins at the transcript and/or protein levels[33]. This observation thus indicates that at least 7 genes (44% of those mostly expressed in microglia) have been already linked to Aβ clearance/toxicity. However, it is also necessary to keep in mind that gene prioritization, while efficient for numerous loci, presents some limitations, particularly in complex locus where it is difficult to clearly identify the most relevant gene. In addition, it is important to note that for our molecular QTL-based analyses, we only considered cis-QTLs that are typically found within 1 Mb around the molecular phenotype feature. Therefore our analyses might have missed certain molecular trans-QTLs with important effect on AD risk. Furthermore, even though we extensively integrated our GWAS results with the information derived from expression, splicing, and methylation landscape of newly identified genetic risk regions that hinted at possible explanation for mechanism of action of the AD association in these regions, our post-GWAS analyses did not account for the possibility that the underlying risk mechanism could be explained through the effect of genetic variation on protein levels (protein QTLs), metabolite levels (metabolite QTLs), 3D spatial organization of chromatin (e.g. topologically associated domains [TADs]) or a structural variant that might be tagged by

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

the newly identified genetic risk variants. We therefore emphasize that for complete elucidation of AD risk mechanisms in these regions, more investigations are required. We therefore cannot exclude that we overstated some pathophysiological pathways based on incorrect gene assignment or incomplete information. Nevertheless, reinforcing the role of microglia in AD, our data also revealed for the first time a (direct or indirect) statistical relationship between gene expression in microglia, genetic risk factors and APP/Aβ pathways (Supplementary table 10).

Translating genetic findings into tools that can be used in the clinical setting has proven to be challenging because different strategies are available to create a PRS summarizing personal genetic burden. In our study, we computed a PRS following the strategy previously described by Chouraki et al[34] in which increase of one point on the PRS corresponds to carrying one additional average risk allele. We observed that this PRS was associated with the risk of progression from MCI to dementia and to AD dementia (ADD), the major form of dementia. Importantly, the association was obtained while the *APOE*-ε4 allele was not included in the PRS. Moreover, the association did not disappear following adjustment for age or additive effect of the *APOE*-ε4 allele. Previous studies evaluating the effect of PRS on progression of MCI to ADD have provided compelling evidence supporting the role of *APOE*, whereas the contribution of additional genetic variants to progression has not shown unanimous results[35]. Our finding of an *APOE*-independent effect on MCI progression may be explained by (i) the large longitudinal sample of 4,609 MCI cases of whom 1,532 converted to dementia (ii) and the improved knowledge of the genetic component of AD through the larger number of newly-discovered genome-wide significant variants included in the PRS. Our study also shows that carrying 18 "average risk variants" conferred a similar probability of progressing to ADD as being *APOE*-ε4 heterozygous, while carrying 32 "average risk variants" resemble that of being *APOE*-ε4 homozygous. Although further research is still needed before these findings can be translated into the clinical routine, completing the genetic architecture of AD is definitely paving the way to personalized risk prediction even before dementia stage is reached.

Of note, several new AD loci were also associated with the risk of developing other neurodegenerative diseases: *IDUA* locus with Parkinson's Disease (PD), *GRN* and *TMEM106B* loci with fronto-temporal dementia (FTD). According to the large number of cases analyzed in our study and the well documented clinical diagnostic errors between neurodegenerative diseases, we cannot exclude that these associations are in part due to contaminations by PD or FTD cases in our sample. However, GWAS colocalization analyses indicate that the main signal in PD is independent of the one observed in AD (Supplementary Tables 25 and 26). Only small GWAS are available in FTD (and depending on the type of FTD)[36], this makes difficult to definitely answer to the independency of the AD and FTD signals. Further investigations will be required all the more since sporadic FTD has been described as a polygenic disorder where multiple pleiotropic loci with small effects contribute to increased disease risk[37]**.** In addition, the lead *GRN* variant in AD is functional and has also been described to be associated with TDP-43 positive FTD risk[38]**.** In this context, it will be interesting to determine whether our PRS may be specific or not of AD given the potential genetic overlap between neurodegenerative diseases. Of note, our PRS was not associated with risk of converting to non-AD dementia (HR=1.010, 95%CI [0.996-1.025], $p=4.8x10^{-1}$). However, this absence of association may be due to lack of statistical power in our current analysis and will require further investigations.

In conclusion, our work demonstrates that improved characterization AD genetics also expands our knowledge of the underlying pathophysiological processes, presenting novel opportunities for therapeutic approaches and risk prediction through robust PRS. Convergence between treatments generated from genomics and PRS may thus pave the way to translational genomics and personalized medicine.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

MATERIALS AND METHODS

Samples. All discovery meta-analysis samples are from the following consortia/datasets: EADB, GR@ACE, EADI, GERAD/PERADES, DemGene, Bonn, the Rotterdam study, the CCHS study, NxC and the UK Biobank. Summary demographics of these case-control studies are described in Supplementary Table 1 and more detailed descriptions are available in the Supplementary Information. Replication samples are from the ADGC and CHARGE consortia (Supplementary Table 1 and Supplementary Information) and fully described elsewhere[6,11,14,15,39–41]. Written informed consent was obtained from study participants or, for those with substantial cognitive impairment, from a caregiver, legal guardian, or other proxy. Study protocols for all cohorts were reviewed and approved by the appropriate institutional review boards. Further details of all cohorts can be found in the Supplementary Information.

Quality control and imputation. Standard quality control was performed on variants and samples on all datasets individually. The samples were then imputed with the Trans-Omics for Precision Medicine (TOPMed) reference panel[42,43]. The Haplotype Reference Consortium (HRC) panel[44] was also used for some datasets (Supplementary Table 2). For the UK Biobank, we used the provided imputed data, generated from a combination of the 1000 Genomes (1000G), HRC and UK10K reference panels. See Supplementary Information for more details.

Stage I analyses. Association tests between AD clinical or proxy status and autosomal genetic variant were conducted separately in each dataset using logistic regression assuming an additive genetic model as implemented in SNPTEST[45] or in PLINK[9], except in the UK Biobank where a logistic mixed model as implemented in SAIGE[46] was considered. Analyses were performed on the genotype probabilities in SNPTEST (newml method) and on dosage in PLINK and SAIGE. Analyses were adjusted for principal components and genotyping centers when necessary (Supplementary Table 2). For the UK Biobank dataset, effect sizes and standard errors were corrected by a factor of two to take into account that proxy cases were analysed[12]. We filtered out duplicated variants and variants with (i) missing effect size, standard error or P value, (ii) absolute value of effect size above 5, (iii) imputation quality less than 0.3, (iv) the product of the minor allele count and the imputation quality (mac-info score) less than 20. In the UK Biobank dataset, only variants with minor allele frequency (MAF) above 0.01% were analyzed. For datasets not imputed with the TOPMed reference panel, we also excluded (i) variants for which conversion of position or alleles from the GRCh37 assembly to the GRCh38 assembly was not possible or problematic, or (ii) variants with very large difference of frequency between the TOPMed reference panel and the reference panels used to perform imputation.
Results were then combined across studies with a fixed-effect meta-analysis using the inverse variance weighted approach as implemented in the METAL software[47]. We filtered (i) variants with heterogeneity P value below $5x10^{-8}$, (ii) variants analyzed in less than 20% of the total number of cases and (iii) variants with frequency amplitude above 0.4 (defined as the difference between the maximum and minimum frequency across studies). We further excluded variants analyzed in the UK Biobank only or variants not analyzed in the EADB-TOPMed dataset.
Genomic inflation factor lambda was computed with the R package GenABEL[48] using the median approach after exclusion of the *APOE* region (44 Mb to 46 Mb on chromosome 19 in GRCh38). The linkage disequilibrium (LD) score (LDSC) regression intercept was computed with the LDSC software using the "baselineLD" LD scores built from 1000 Genomes Phase 3 [8]. The analysis was restricted to HapMap 3 variants and excluded multi-allelic variants, variants without an rsID and variants in the *APOE* region.

Definition of associated loci. A region of +/- 500kb was defined around each variant with a Stage I P value below $1x10^{-5}$. Those regions were then merged with the bedtools software to define non-overlapping regions. The region corresponding to the APOE locus was excluded.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

We then applied the PLINK clumping procedure to define independent hits in each of those regions. The clumping procedure was applied on all variants with a Stage I P value below $1x10^{-5}$. It is an iterative process beginning with the variant with the lowest P value, named index variant. Variants with a Stage I P value below $1x10^{-5}$, located within 500 kb of this index variant, and in LD with the index variant (r^2 above 0.001) are assigned to the clump of this index variant. The clumping procedure is then applied on all the remaining variants, until no variant is left. LD was computed in the EADB-TOPMed dataset using high quality (probability ≥ 0.8) imputed genotypes.

Stage II analyses. Variants with a Stage I P value below $1x10^{-5}$ were sent for follow-up (see Supplementary Information). A fixed-effect meta-analysis was performed with METAL (inverse variance weighted approach) to combine the results across Stage I and Stage II. In each clump, we then reported the replicated variant (same direction of effects between Stage I and Stage II, with a Stage II P value below 0.05) with the lowest P value in the meta-analysis of the Stages I + II. Those variants were considered associated at the genome-wide significance level if they had a P value below $5x10^{-8}$ in the Stages I + II meta-analysis. Among them, we excluded the variant chr6:32657066:G:A because its frequency amplitude was large.

Pathway analysis. The assignment of Gene Ontology (GO) terms to human genes was obtained from the "gene2go" file, downloaded from NCBI on March 11th 2020. "Parent" GO terms were assigned to genes using the ontology file downloaded from the Gene Ontology website on the same date. GO terms were assigned to genes based on experimental or curated evidence of a specific type, so evidence codes IEA (electronic annotation), NAS (non-traceable author statement), RCA (inferred from reviewed computational analysis) were excluded. Pathways were downloaded from the Reactome website on April 26th 2020. Biocarta, KEGG and Pathway Interaction Database (PID) pathways were downloaded from v7.1 (March 2020) of the Molecular Signatures Database. Analysis was restricted to GO terms containing between 10 and 2000 genes. No size restrictions were placed on the other gene sets, since there were many fewer of them. This resulted in a total of 10,271 gene sets for analysis. Gene set enrichment analyses were performed in MAGMA[49], correcting for the number of variants in each gene, linkage disequilibrium (LD) between variants and LD between genes. LD was computed in the EADB-TOPMed dataset using high quality (probability ≥ 0.9) imputed genotypes. The measure of pathway enrichment is the MAGMA "competitive" test (where the association statistic for genes in the pathway is compared to those of all other protein-coding genes), as recommended by De Leeuw et al.[50]. We used the "mean" test statistic, which uses the sum of -log(variant P value) across all genes as the association statistic for genes. The primary analysis assigned variants to genes if they lie within the gene boundaries, but a secondary analysis used a window of 35kb upstream and 10kb downstream to assign variants to genes, as in Kunkle et al[6]. The primary analysis used all variants with imputation quality above 0.8. We used q-values[51] to account for multiple testing throughout this report.

QTLs/TWAS/MetaMeth. In order to prioritize candidate genes in the new loci, we employed several approaches: (i) expression quantitative trait loci (eQTLs) and colocalization (eQTL coloc) analyses combined with expression transcriptome-wide association studies (eTWAS) in AD-relevant brain regions; (ii) splicing quantitative trait loci (sQTL) and colocalization (sQTL coloc) analyses combined with splicing transcriptome-wide association studies (sTWAS) in AD-relevant brain regions; (iii) genetic-driven methylation as a biological mediator of genetic signals in blood (MetaMeth). In our regions of interest, we systematically searched if a gene has a significant e/sQTL, colocalization e/sTWAS and/or MetaMeth signal(s) within a region of 1 Mb around the lead variant. In addition to the "nearest" genes from the lead variant, we kept for further analyses those exhibiting such AD-related modulations. We then added several additional approaches: (i) data from a genome-wide, high-content siRNA screening approach to assess the functional impact of gene under-

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

expression on APP metabolism[20]; (ii) methylation QTL (mQTL) and histone acetylation QTL (haQTL) effects of the lead variants in DLPFC[21] and (iii) additional eQTL effects of these variants in monocytes and macrophages[22–27]. A full description of how the genes were prioritized is reported in the Supplementary Information (see also Supplementary Tables 12-21 and Supplementary Fig. 26-34).

Cell type expression. Assignment of newly identified AD risk genes to specific cell classes of the adult brain was performed as previously described[52]. Briefly, middle temporal gyrus (MTG) single-nucleus transcriptomes (15,928 total nuclei derived from 8 human tissue donors ranging in age from 24-66 year), were used to annotate and select 6 main cell classes using S*eurat 3.1.1*[53]: Glutamatergic Neurons, GABAergic Neurons, Astrocytes, Oligodendrocytes, Microglia and Endothelial cells.

PRS analysis. Twelve longitudinal MCI cohorts were included in the analysis and are fully described in the Supplementary Information and Supplementary Table 23.
PRS were calculated as previously described[34]. Briefly, we considered 60 variants with genome-wide significant evidence of association with AD in our study (Figure 1, Tables 1 and 2) with a MAF ≥ 0.05 in our case-control study. Variants were directly genotyped or imputed ($R^2 \geq 0.3$). We did not include any *APOE* variants in the PRS. The PRS was calculated as the weighted average of the number of risk increasing alleles for each variant. Weights were based on the respective log(OR) obtained in the Stage II since no samples in this stage were included in the MCI study. The PRS was then multiplied by 60, i.e. the number of included variants. Thus, an increase in HR corresponds to carry one additional average risk allele.
All PCs used were generated per cohort, using the same variants that were used on the case/control study PCA. The number of *APOE*-e4 alleles was obtained based on direct genotyping or, if missing, based on genotypes derived from the TOPMed imputations.
The association of the PRS with risk of progression to dementia in patients with MCI was assessed using Cox proportional hazards-regression analysis. First, the effect was analyzed on progression to all-cause dementia (i.e. regardless of clinical dementia subtype). Next, the analysis was focused on MCI patients converting to AD dementia. To this end, all converters to non-AD dementia were excluded from the analysis sample. Two cohorts (HBA, SAN) were excluded due to missing information on the clinical dementia subtype at this stage. Finally, to assess whether the exclusion of non-AD dementia converters affected our results, the analysis was repeated by coding non-AD converters as censored cases.
Each Cox-regression analysis was first performed unadjusted for covariates and then repeated, adjusted for age, sex and the first four principal components to correct for potential population stratification. Furthermore, analyses were additionally controlled for the number of *APOE*-e4 alleles (assuming an additive effect) to assess the independence of the PRS effect from *APOE*. Moreover, the interaction between the PRS and *APOE*-e4 was tested.

URLs:

Bedtools: https://bedtools.readthedocs.io/en/latest/
BCFtools: http://samtools.github.io/bcftools/bcftools.html
Samtools: http:/www.htslib.org/doc/samtools.html
gene2go: ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/
Gene Ontology: http://geneontology.org/docs/download-ontology/
Reactome: https://reactome.org/download-data
KEGG and Pathway Interaction Database (PID) pathways: https://www.gsea-msigdb.org/gsea/msigdb/index.jsp
AMP-AD rnaSeqReprocessing Study: https://www.synapse.org/#!Synapse:syn9702085
MayoRNAseq WGS VCFs: https://www.synapse.org/#!Synapse:syn11724002
ROSMAP WGS VCFs: https://www.synapse.org/#!Synapse:syn11724057
MSBB WGS VCFs: https://www.synapse.org/#!Synapse:syn11723899
GTEx pipeline: https://github.com/broadinstitute/gtex-pipeline

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Leafcutter: https://github.com/davidaknowles/leafcutter
RegTools: https://github.com/griffithlab/regtools
Enhanced version of FastQTL: https://github.com/francois-a/fastqtl
Picard: https://broadinstitute.github.io/picard/
eQTLGen: https://www.eqtlgen.org/
eQTL Catalogue database: https://www.ebi.ac.uk/eqtl/
Brain xQTL serve: http://mostafavilab.stat.ubc.ca/xqtl/
GTEx v8 eQTL and sQTL catalogues: https://www.gtexportal.org/
coloc: https://github.com/chr1swallace/coloc
FUSION: https://github.com/gusevlab/fusion_twas
GTEx v8 expression and splicing prediction models: http://predictdb.org/
MetaXcan: https://github.com/hakyimlab/MetaXcan
FOCUS: https://github.com/bogdanlab/focus
qcat: https://github.com/nanoporetech/qcat
minimap2: https://github.com/lh3/minimap2
NanoStat: https://github.com/wdecoster/nanostat
mosdepth: https://github.com/brentp/mosdepth
ggplot2: https://ggplot2.tidyverse.org/
LocusZoom: https://github.com/statgen/locuszoom-standalone
pyGenomeTracks: https://github.com/deeptools/pyGenomeTracks
VCFs of phased biallelic SNV and INDEL variants of 1KG samples (de novo called on GRCh38):
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20190312_biallelic_SNV_and_INDEL/

References:

1. Zeggini, E., Gloyn, A. L., Barton, A. C. & Wain, L. V. Translational genomics and precision medicine: Moving from the lab to the clinic. *Science (80-.).* **365**, 1409–1413 (2019).

2. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–74 (2006).

3. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* **41**, 1088–93 (2009).

4. Lambert, J.-C. *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.* **41**, 1094–9 (2009).

5. Bellenguez, C., Grenier-Boley, B. & Lambert, J. C. Genetics of Alzheimer's disease: where we are, and where we are going. *Current Opinion in Neurobiology* vol. 61 40–48 (2020).

6. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).

7. Kowalski, M. H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet.* **15**, (2019).

8. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

9. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

10. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

11. 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).

12. 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).

13. Marioni, R. E. *et al.* GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* **8**, (2018).

14. Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* (2017) doi:10.1038/ng.3916.

15. Jun, G. *et al.* A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol. Psychiatry* **21**, 108–17 (2016).

16. Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping, and integrative prioritization identify new Alzheimer's disease risk genes. *medRxiv* 2020.01.22.20018424 (2020) doi:10.1101/2020.01.22.20018424.

17. de Rojas, I. *et al.* Common variants in Alzheimer's disease: Novel association of six genetic variants with AD and risk stratification by polygenic risk scores. *medRxiv* 19012021 (2019) doi:10.1101/19012021.

18. Escott-Price, V. *et al.* Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. *PLoS One* **9**, e94661 (2014).

19. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).

20. 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).

21. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).

22. Alasoo, K. *et al.* Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat. Genet.* **50**, 424–431 (2018).

23. Nédélec, Y. *et al.* Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **167**, 657-669.e21 (2016).

24. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398-1414.e24 (2016).

25. Momozawa, Y. *et al.* IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, (2018).

26. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science (80-.).* **343**, (2014).

27. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**, 643-656.e17 (2016).

28. Holstege, H. *et al.* Exome sequencing identifies novel AD-associated genes. *medRxiv* (2020) doi:10.1101/2020.07.22.20159251.

29. Kim, T. *et al.* Human LilrB2 is a β-amyloid receptor and its murine homolog PirB regulates synaptic plasticity in an Alzheimer's model. *Science (80-.).* **341**, 1399–1404 (2013).

30. Cao, Q. *et al.* Inhibiting amyloid-ß cytotoxicity through its interaction with the cell surface receptor LilrB2 by structure-based design. *Nat. Chem.* **10**, (2018).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

31. Salminen, A. & Kaarniranta, K. Siglec receptors and hiding plaques in Alzheimer's disease. *Journal of Molecular Medicine* vol. 87 697–701 (2009).

32. Takahashi, H. *et al.* Opposing effects of progranulin deficiency on amyloid and tau pathologies via microglial TYROBP network. *Acta Neuropathol.* **133**, 785–807 (2017).

33. Sierksma, A. *et al.* Novel Alzheimer risk genes determine the microglia response to amyloid-β but not to TAU pathology. *EMBO Mol. Med.* **12**, (2020).

34. Chouraki, V. *et al.* Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease. *J. Alzheimers. Dis.* **53**, 921–32 (2016).

35. Harrison, J. R., Mistry, S., Muskett, N., Escott-Price, V. & Brookes, K. From Polygenic Scores to Precision Medicine in Alzheimer's Disease: A Systematic Review. *Journal of Alzheimer's Disease* vol. 74 1271–1283 (2020).

36. Ferrari, R. *et al.* Frontotemporal dementia and its subtypes: A genome-wide association study. *Lancet Neurol.* **13**, 686–699 (2014).

37. Ferrari, R. *et al.* Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *J. Neurol. Neurosurg. Psychiatry* **88**, 152–164 (2017).

38. Rademakers, R. *et al.* Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia. *Hum. Mol. Genet.* **17**, 3631–3642 (2008).

39. Psaty, B. M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium design of prospective meta-analyses of genome-wide association studies from 5 Cohorts. *Circulation: Cardiovascular Genetics* vol. 2 73–80 (2009).

40. 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**, 436–41 (2011).

41. 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**, 1473–84 (2010).

42. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv* **2**, 563866 (2019).

43. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

44. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–83 (2016).

45. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–13 (2007).

46. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

47. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–1 (2010).

48. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).

49. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, (2015).

50. De Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016).

51. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad.*

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Sci. U. S. A. **100**, 9440–9445 (2003).

52. Coelho, D. M., Carvalho, L. I. da C., Melo-de-Farias, A. R., Lambert, J.-C. & Costa, M. R. Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains. *medRxiv* 2020.03.19.20038703 (2020) doi:10.1101/2020.03.19.20038703.

53. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Acknowledgment

EADB. We thank the numerous participants, researchers, and staff from many studies who collected and contributed to the data. We thank the informatics resource centre of the University of Lille. This research has been conducted using the UK Biobank Resource *under Application Number* 61054.

This work was supported by a grant (European Alzheimer DNA BioBank, EADB) from the EU Joint Programme – Neurodegenerative Disease Research (JPND). Inserm UMR1167 is also funded by Inserm, Institut Pasteur de Lille, the Lille Métropole Communauté Urbaine, the French government's LABEX DISTALZ program (development of innovative strategies for a transdisciplinary approach to Alzheimer's disease).

Additional support for EADB cohorts was provided by: Italian Ministry of Health (Ricerca Corrente); Ministero dell'Istruzione, dell'Università e della Ricerca–MIUR project "Dipartimenti di Eccellenza 2018–2022" to Department of Neuroscience "Rita Levi Montalcini", University of Torino (IR), and AIRAlzh Onlus-ANCC-COOP (SB); Partly supported by "Ministero della Salute", I.R.C.C.S. Research Program, Ricerca Corrente 2018-2020, Linea n. 2 "Meccanismi genetici, predizione e terapie innovative delle malattie complesse" and by the "5 x 1000" voluntary contribution to the Fondazione I.R.C.C.S. Ospedale "Casa Sollievo della Sofferenza"; and RF-2018-12366665, Fondi per la ricerca 2019 (Sandro Sorbi). Copenhagen General Population Study (CGPS): We thank staff and participants of the CGPS for their important contributions. Karolinska Institutet AD cohort: Dr. Graff and co-authors of the Karolinska Institutet AD cohort report grants from Swedish Research Council (VR) 2015-02926, 2018-02754, 2015-06799, Swedish Alzheimer Foundation, Stockholm County Council ALF and resarch school, Karolinska Institutet StratNeuro, Swedish Demensfonden, and Swedish brain foundation, during the conduct of the study. ADGEN: This work was supported by Academy of Finland (grant numbers 307866); Sigrid Jusélius Foundation; the Strategic Neuroscience Funding of the University of Eastern Finland; EADB project in the JPNDCO-FUND program (grant number 301220). CBAS: Supported by the project no. LQ1605 from the National Program of Sustainability II (MEYS CR), Supported by Ministry of Health of the Czech Republic, grant nr. NV19-04-00270 (All rights reserved), Grant Agency of Charles University Grants No. 693018 and 654217; the Ministry of Health, Czech Republic—conceptual development of research organization, University Hospital Motol, Prague, Czech Republic Grant No. 00064203; the Czech Ministry of Health Project AZV Grant No. 16—27611A; and Institutional Support of Excellence 2. LF UK Grant No. 699012. CNRMAJ-Rouen: This study received fundings from the Centre National de Référence Malades Alzheimer Jeunes (CNRMAJ). The Finnish Geriatric Intervention Study for the Prevention of Cognitive Impairment and Disability (FINGER) data collection was supported by grants from the Academy of Finland, La Carita Foundation, Juho Vainio Foundation, Novo Nordisk Foundation, Finnish Social Insurance Institution, Ministry of Education and Culture Research Grants, Yrjö Jahnsson Foundation, Finnish Cultural Foundation South Osthrobothnia Regional Fund, and EVO/State Research Funding grants of University Hospitals of Kuopio, Oulu and Turku, Seinäjoki Central Hospital and Oulu City Hospital, Alzheimer's Research & Prevention Foundation USA, AXA Research Fund, Knut and Alice Wallenberg Foundation Sweden, Center for Innovative Medicine (CIMED) at Karolinska Institutet Sweden, and Stiftelsen Stockholms sjukhem Sweden. FINGER cohort genotyping was funded by EADB project in the JPND CO-FUND (grant number 301220). Research on the Belgian EADB cohort is funded in part by the Alzheimer Research Foundation (SAO-FRA), The Research Foundation Flanders (FWO), and the University of Antwerp Research Fund. FK is supported by a BOF DOCPRO fellowship of the University of Antwerp Research Fund. SNAC-K is financially supported by the Swedish Ministry of Health and Social Affairs, the participating County Councils and Municipalities, and the Swedish Research Council. BDR Bristol: We would like to thank the South West Dementia Brain Bank (SWDBB) for

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

providing brain tissue for this study. The SWDBB is part of the Brains for Dementia Research programme, jointly funded by Alzheimer's Research UK and Alzheimer's Society and is supported by BRACE (Bristol Research into Alzheimer's and Care of the Elderly) and the Medical Research Council. BDR Manchester: We would like to thank the Manchester Brain Bankfor providing brain tissue for this study. The Manchester Brain Bank is part of the Brains for Dementia Research programme, jointly funded by Alzheimer's Research UK and Alzheimer's Society. BDR KCL: Human post-mortem tissue was provided by the London Neurodegenerative Diseases Brain Bank which receives funding from the UK Medical Research Council and as part of the Brains for Dementia Research programme, jointly funded by Alzheimer's Research UK and the Alzheimer's Society. The CFAS Wales study was funded by the ESRC (RES-060-25-0060) and HEFCW as 'Maintaining function and well-being in later life: a longitudinal cohort study', (Principal Investigators: R.T Woods, L.Clare, G.Windle, V. Burholt, J. Philips, C. Brayne, C. McCracken, K. Bennett, F. Matthews). We are grateful to the NISCHR Clinical Research Centre for their assistance in tracing participants and in interviewing and in collecting blood samples, and to general practices in the study areas for their cooperation. MRC: We thank all individuals who participated in this study. Cardiff University was supported by the Alzheimer's Society (AS; grant RF014/164) and the Medical Research Council (MRC; grants G0801418/1, MR/K013041/1, MR/L023784/1) (R. Sims is an AS Research Fellow). Cardiff University was also supported by the European Joint Programme for Neurodegenerative Disease (JPND; grant MR/L501517/1), Alzheimer's Research UK (ARUK; grant ARUK-PG2014-1), the Welsh Assembly Government (grant SGR544:CADR), Brain's for dementia Research and a donation from the Moondance Charitable Foundation. Cardiff University acknowledges the support of the UK Dementia Research Institute, of which J. Williams is an associate director. Cambridge University acknowledges support from the MRC. Patient recruitment for the MRC Prion Unit/UCL Department of Neurodegenerative Disease collection was supported by the UCLH/UCL Biomedical Centre and NIHR Queen Square Dementia Biomedical Research Unit. The University of Southampton acknowledges support from the AS. King's College London was supported by the NIHR Biomedical Research Centre for Mental Health and the Biomedical Research Unit for Dementia at the South London and Maudsley NHS Foundation Trust and by King's College London and the MRC. ARUK and the Big Lottery Fund provided support to Nottingham University. Alfredo Ramirez: Part of the work was funded by the JPND EADB grant (German Federal Ministry of Education and Research (BMBF) grant: 01ED1619A). German Study on Ageing, Cognition and Dementia in Primary Care Patients (AgeCoDe): This study/publication is part of the German Research Network on Dementia (KND), the German Research Network on Degenerative Dementia (KNDD; German Study on Ageing, Cognition and Dementia in Primary Care Patients; AgeCoDe), and the Health Service Research Initiative (Study on Needs, health service use, costs and health-related quality of life in a large sample of oldestold primary care patients (85+; AgeQualiDe)) and was funded by the German Federal Ministry of Education and Research (grants KND: 01GI0102, 01GI0420, 01GI0422, 01GI0423, 01GI0429, 01GI0431, 01GI0433, 01GI0434; grants KNDD: 01GI0710, 01GI0711, 01GI0712, 01GI0713, 01GI0714, 01GI0715, 01GI0716; grants Health Service Research Initiative: 01GY1322A, 01GY1322B, 01GY1322C, 01GY1322D, 01GY1322E, 01GY1322F, 01GY1322G). VITA study: The support of the Ludwig Boltzmann Society and the AFI Germany have supported the VITA study. The former VITA study group should be acknowledged: W. Danielczyk, G. Gatterer, K Jellinger, S Jugwirth, KH Tragl, S Zehetmayer. Vogel Study: This work was financed by a research grant of the ''Vogelstiftung Dr. Eckernkamp''. HELIAD study: This study was supported by the grants: IIRG-09-133014 from the Alzheimer's Association, 189 10276/8/9/2011 from the ESPA-EU program Excellence Grant (ARISTEIA) and the ΔΥ2β/οικ.51657/14.4.2009 of the Ministry for Health and Social Solidarity (Greece). Biobank Department of Psychiatry, UMG: Prof. Jens Wiltfang is supported by an Ilídio Pinho professorship and iBiMED (UID/BIM/04501/2013), and FCT project PTDC/DTP_PIC/5587/2014 at the University of Aveiro, Portugal. Lausanne study: This work was supported by grants from the Swiss National Research Foundation (SNF 320030_141179). PAGES study: Harald Hampel is an employee of Eisai Inc. During part of

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

this work he was supported by the AXA Research Fund, the "Fondation partenariale Sorbonne Université" and the "Fondation pour la Recherche sur Alzheimer", Paris, France. Mannheim, Germany Biobank: Department of geriatric Psychiatry, Central Institute for Mental Health, Mannheim, University of Heidelberg, Germany. Genotyping for the Swedish Twin Studies of Aging was supported by NIH/NIA grant R01 AG037985. Genotyping in TwinGene was supported by NIH/NIDDK U01 DK066134. WvdF is recipient of Joint Programming for Neurodegenerative Diseases (JPND) grants PERADES (ANR-13-JPRF-0001) and EADB (733051061). Gothenburg Birth Cohort (GBC) Studies: We would like to thank UCL Genomics for performing the genotyping analyses. The studies were supported by The Stena Foundation, The Swedish Research Council (2015-02830, 2013-8717), The Swedish Research Council for Health, Working Life and Wellfare (2013-1202, 2005-0762, 2008-1210, 2013-2300, 2013- 2496, 2013-0475), The Brain Foundation, Sahlgrenska University Hospital (ALF), The Alzheimer's Association (IIRG-03-6168), The Alzheimer's Association Zenith Award (ZEN-01-3151), Eivind och Elsa K:son Sylvans Stiftelse, The Swedish Alzheimer Foundation. Clinical AD, Sweden: We would like to thank UCL Genomics for performing the genotyping analyses. Barcelona Brain Biobank: Brain Donors of the Neurological Tissue Bank of the Biobanc-Hospital Clinic-IDIBAPS and their families for their generosity. Hospital Clínic de Barcelona Spanish Ministry of Economy and Competitiveness-Instituto de Salud Carlos III and Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea, "Una manera de hacer Europa" grants (PI16/0235 to Dr. R. Sánchez-Valle and PI17/00670 to Dr. A.Antonell). AA is funded by Departament de Salut de la Generalitat de Catalunya, PERIS 2016-2020 (SLT002/16/00329). Sydney Memory and Ageing Study (Sydney MAS): We gratefully acknowledge and thank the following for their contributions to Sydney MAS: participants, their supporters and the Sydney MAS Research Team (current and former staff and students). Funding was awarded from the Australian National Health and Medical Research Council (NHMRC) Program Grants (350833, 568969, 109308). AddNeuroMed consortium was led by Simon Lovestone, Bruno Vellas, Patrizia Mecocci, Magda Tsolaki, Iwona Kłoszewska, Hilkka Soininen. This work was supported by InnoMed (Innovative Medicines in Europe), an integrated project funded by the European Union of the Sixth Framework program priority (FP6-2004- LIFESCIHEALTH-5). Oviedo: This work was partly supported by Grant from Fondo de Investigaciones Sanitarias-Fondos FEDER EuropeanUnion to Victoria Alvarez PI15/00878. Pascual Sánchez-Juan is supported by CIBERNED and Carlos III Institute of Health, Spain (PI08/0139, PI12/02288, and PI16/01652), jointly funded by Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea, "Una manera de hacer Europa". Project MinE: The ProjectMinE study was supported by the ALS Foundation Netherlands and the MND association (UK) (Project MinE, www.projectmine.com). The SPIN cohort: We are indebted to patients and their families for their participation in the "Sant Pau Initiative on Neurodegeneration cohort", at the Sant Pau Hospital (Barcelona). This is a multimodal research cohort for biomarker discovery and validation that is partially funded by Generalitat de Catalunya (2017 SGR 547 to JC), as well as from the Institute of Health Carlos III-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER- "Una manera de Hacer Europa") (grants PI11/02526, PI14/01126, and PI17/01019 to JF; PI17/01895 to AL), and the Centro de Investigación Biomédica en Red Enfermedades Neurodegenerativas programme (Program 1, Alzheimer Disease to AL). We would also like to thank the Fundació Bancària Obra Social La Caixa (DABNI project) to JF and AL; and Fundación BBVA (to AL), for their support in funding this follow-up study. Adolfo López de Munain is supported by Fundación Salud 2000 (PI2013156) and Diputación Foral de Gipuzkoa (Exp.114/17).

Gra@ce. The Genome Research @ Fundació ACE project (GR@ACE) is supported by Grifols SA, Fundación bancaria 'La Caixa', Fundació ACE, and CIBERNED. A.R. and M.B. receive support from the European Union/EFPIA Innovative Medicines Initiative Joint undertaking ADAPTED and MOPEAD projects (grant numbers 115975 and 115985, respectively). M.B. and A.R. are also supported by national grants PI13/02434, PI16/01861, PI17/01474 and PI19/01240. Acción Estratégica en Salud is integrated into the Spanish National R + D + I Plan and funded by ISCIII (Instituto de Salud Carlos III)–Subdirección

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER–'Una manera de hacer Europa'). Some control samples and data from patients included in this study were provided in part by the National DNA Bank Carlos III (www.bancoadn.org, University of Salamanca, Spain) and Hospital Universitario Virgen de Valme (Sevilla, Spain); they were processed following standard operating procedures with the appropriate approval of the Ethical and Scientific Committee. The present work has been performed as part of the doctoral program of I. de Rojas at the Universitat de Barcelona (Barcelona, Spain).
EADI. This work has been developed and supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant (Development of Innovative Strategies for a Transdisciplinary approach to ALZheimer's disease) including funding from MEL (Metropole européenne de Lille), ERDF (European Regional Development Fund) and Conseil Régional Nord Pas de Calais. This work was supported by INSERM, the National Foundation for Alzheimer's disease and related disorders, the Institut Pasteur de Lille and the Centre National de Recherche en Génomique Humaine, CEA, the JPND PERADES, the Laboratory of Excellence GENMED (Medical Genomics) grant no. ANR-10-LABX-0013 managed by the National Research Agency (ANR) part of the Investment for the Future program, and the FP7 AgedBrainSysBio. The Three-City Study was performed as part of collaboration between the Institut National de la Santé et de la Recherche Médicale (Inserm), the Victor Segalen Bordeaux II University and Sanofi-Synthélabo. The Fondation pour la Recherche Médicale funded the preparation and initiation of the study. The 3C Study was also funded by the Caisse Nationale Maladie des Travailleurs Salariés, Direction Générale de la Santé, MGEN, Institut de la Longévité, Agence Française de Sécurité Sanitaire des Produits de Santé, the Aquitaine and Bourgogne Regional Councils, Agence Nationale de la Recherche, ANR supported the COGINUT and COVADIS projects. Fondation de France and the joint French Ministry of Research/INSERM "Cohortes et collections de données biologiques" programme. Lille Génopôle received an unconditional grant from Eisai. The Three-city biological bank was developed and maintained by the laboratory for genomic analysis LAG-BRC - Institut Pasteur de Lille.

GERAD/PERADES. We thank all individuals who participated in this study. Cardiff University was supported by the Wellcome Trust, Alzheimer's Society (AS; grant RF014/164), the Medical Research Council (MRC; grants G0801418/1, MR/K013041/1, MR/L023784/1), the European Joint Programme for Neurodegenerative Disease (JPND, grant MR/L501517/1), Alzheimer's Research UK (ARUK, grant ARUK-PG2014-1), Welsh Assembly Government (grant SGR544:CADR), a donation from the Moondance Charitable Foundation, UK Dementia's Platform (DPUK, reference MR/L023784/1), and the UK Dementia Research Institute at Cardiff. Cambridge University acknowledges support from the MRC. ARUK supported sample collections at the Kings College London, the South West Dementia Bank, Universities of Cambridge, Nottingham, Manchester and Belfast. King's College London was supported by the NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at the South London and Maudsley NHS Foundation Trust and Kings College London and the MRC. Alzheimer's Research UK (ARUK) and the Big Lottery Fund provided support to Nottingham University. Ulster Garden Villages, AS, ARUK, American Federation for Aging Research, NI R&D Office and the Royal College of Physicians/Dunhill Medical Trust provided support for Queen's University, Belfast. The University of Southampton acknowledges support from the AS. The MRC and Mercer's Institute for Research on Ageing supported the Trinity College group. DCR is a Wellcome Trust Principal Research fellow. The South West Dementia Brain Bank acknowledges support from Bristol Research into Alzheimer's and Care of the Elderly. The Charles Wolfson Charitable Trust supported the OPTIMA group. Washington University was funded by NIH grants, Barnes Jewish Foundation and the Charles and Joanne Knight Alzheimer's Research Initiative. Patient recruitment for the MRC Prion Unit/UCL Department of Neurodegenerative Disease collection was supported by the UCLH/UCL Biomedical Research Centre and their work was supported by the NIHR Queen Square Dementia BRU, the Alzheimer's Research UK and the Alzheimer's Society. LASER-AD was funded by Lundbeck SA. The AgeCoDe study group was supported by the

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

German Federal Ministry for Education and Research grants 01 GI 0710, 01 GI 0712, 01 GI 0713, 01 GI 0714, 01 GI 0715, 01 GI 0716, 01 GI 0717. Genotyping of the Bonn case-control sample was funded by the German centre for Neurodegenerative Diseases (DZNE), Germany. The GERAD Consortium also used samples ascertained by the NIMH AD Genetics Initiative. HH was supported by a grant of the Katharina-Hardt-Foundation, Bad Homburg vor der Höhe, Germany. The KORA F4 studies were financed by Helmholtz Zentrum München; German Research Center for Environmental Health; BMBF; German National Genome Research Network and the Munich Center of Health Sciences. The Heinz Nixdorf Recall cohort was funded by the Heinz Nixdorf Foundation (Dr. Jur. G.Schmidt, Chairman) and BMBF. We acknowledge use of genotype data from the 1958 Birth Cohort collection and National Blood Service, funded by the MRC and the Wellcome Trust which was genotyped by the Wellcome Trust Case Control Consortium and the Type-1 Diabetes Genetics Consortium, sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development and Juvenile Diabetes Research Foundation International. The project is also supported through the following funding organisations under the aegis of JPND - www.jpnd.eu (United Kingdom, Medical Research Council (MR/L501529/1; MR/R024804/1) and Economic and Social Research Council (ES/L008238/1)) and through the Motor Neurone Disease Association. This study represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. Prof Jens Wiltfang is supported by an Ilídio Pinho professorship and iBiMED (UID/BIM/04501/2013), at the University of Aveiro, Portugal.

Rotterdam study. Rotterdam (RS). This study was funded by the Netherlands Organisation for Health Research and Development (ZonMW) as part of the Joint Programming for Neurological Disease (JPND)as part of the PERADES Program (Defining Genetic Polygenic, and Environmental Risk for Alzheimer's disease using multiple powerful cohorts, focused Epigenetics and Stem cell metabolomics), Project number 733051021. This work was funded also by the European Union Innovative Medicine Initiative (IMI) programme under grant agreement No. 115975 as part of the Alzheimer's Disease Apolipoprotein Pathology for Treatment Elucidation and Development (ADAPTED, https://www.imi-adapted.eu);and the European Union's Horizon 2020 research and innovation programme as part of the Common mechanisms and pathways in Stroke and Alzheimer's disease CoSTREAM project (www.costream.eu, grant agreement No. 667375). The current study is supported by the Deltaplan Dementie and Memorabel supported by ZonMW (Project number 733050814) and Alzheimer Nederland. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of GWAS genotype data for the Rotterdam Study (RS-I, RS-II, RS-III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organization of Scientific Research NWO Investments (Project number 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project number 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina- Gomez, MSc, for the creation and analysis of imputed data.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

DemGene. The project has received funding from The Research Council of Norway (RCN) Grant Nos. 213837, 223273, 225989, 248778, and 251134 and EU JPND Program ApGeM RCN Grant No. 237250, the South-East Norway Health Authority Grant No. 2013-123, the Norwegian Health Association, and KG Jebsen Foundation. The RCN FRIPRO Mobility grant scheme (FRICON) is co-funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under Marie Curie grant agreement No 608695. European Community's grant PIAPP-GA-2011-286213 PsychDPC.

Bonn study. This group would like to thank Dr. Heike Koelsch for her scientific support. The Bonn group was funded by the German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant number 01GI0102, 01GI0711, 01GI042

ADGC. The National Institutes of Health, National Institute on Aging (NIH-NIA) supported this work through the following grants: ADGC, U01 AG032984, RC2 AG036528; Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible; Data for this study were prepared, archived, and distributed by the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689-01); NACC, U01 AG016976; NIA LOAD (Columbia University), U24 AG026395, U24 AG026390, R01AG041797; Banner Sun Health Research Institute P30 AG019610; Boston University, P30 AG013846, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG017173, R01 AG025259, R01 AG048927, R01AG33193, R01 AG009029; Columbia University, P50 AG008702, R37 AG015473, R01 AG037212, R01 AG028786; Duke University, P30 AG028377, AG05128; Einstein Aging Study NIA grant at Albert Einstein College of Medicine, P01 AG03949. Emory University, AG025688; Group Health Research Institute, UO1 AG006781, UO1 HG004610, UO1 HG006375, U01 HG008657; Indiana University, P30 AG10133, R01 AG009956, RC2 AG036650; Johns Hopkins University, P50 AG005146, R01 AG020688; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG016574, R01 AG032990, KL2 RR024151; Mount Sinai School of Medicine, P50 AG005138, P01 AG002219; New York University, P30 AG08051, UL1 RR029893, 5R01AG012101, 5R01AG022374, 5R01AG013616, 1RC2AG036502, 1R01AG035137; North Carolina A&T University, P20 MD000546, R01 AG28786-01A1; Northwestern University, P30 AG013854; Oregon Health & Science University, P30 AG008017, R01 AG026916; Rush University, P30 AG010161, R01 AG019085, R01 AG15819, R01 AG17917, R01 AG030146, R01 AG01101, RC2 AG036650, R01 AG22018; TGen, R01 NS059873; University of Alabama at Birmingham, P50 AG016582; University of Arizona, R01 AG031581; University of California, Davis, P30 AG010129; University of California, Irvine, P50 AG016573; University of California, Los Angeles, P50 AG016570; University of California, San Diego, P50 AG005131; University of California, San Francisco, P50 AG023501, P01 AG019724; University of Kentucky, P30 AG028383, AG05144; University of Michigan, P30 AG053760 and AG063760; University of Pennsylvania, P30 AG010124; University of Pittsburgh, P50 AG005133, AG030653, AG041718, AG07562, AG02365; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG012300; University of Miami, R01 AG027944, AG010491, AG027944, AG021547, AG019757; University of Washington, P50 AG005136, R01 AG042437; University of Wisconsin, P50 AG033514; Vanderbilt University, R01 AG019085; and Washington University, P50 AG005681, P01 AG03991, P01 AG026276. HP was supported by AG025711. ER was supported by CCNA. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is funded by NINDS grant # NS39764, NIMH MH60451 and by Glaxo Smith Kline. Support was also from the Alzheimer's Association (LAF, IIRG-08-89720; MP-V, IIRG-05- 14147), the US Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program, and BrightFocus Foundation (MP-V, A2111048). P.S.G.-H. is supported by

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institute of Health Research. Genotyping of the TGEN2 cohort was supported by Kronos Science. The TGen series was also funded by NIA grant AG041232 to AJM and MJH, The Banner Alzheimer's Foundation, The Johnnie B. Byrd Sr. Alzheimer's Institute, the Medical Research Council, and the state of Arizona and also includes samples from the following sites: Newcastle Brain Tissue Resource (funding via the Medical Research Council, local NHS trusts and Newcastle University), MRC London Brain Bank for Neurodegenerative Diseases (funding via the Medical Research Council),South West Dementia Brain Bank (funding via numerous sources including the Higher Education Funding Council for England (HEFCE), Alzheimer's Research Trust (ART), BRACE as well as North Bristol NHS Trust Research and Innovation department and DeNDRoN), The Netherlands Brain Bank (funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stiching Alzheimer Onderzoek), Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona. ADNI data collection and sharing was funded by the National Institutes of Health Grant U01 AG024904 and Department of Defense award number W81XWH-12-2-0012. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. We thank Drs. D. Stephen Snyder and Marilyn Miller from NIA who are *ex-officio* ADGC members.

Charge. Cardiovascular Health Study (CHS). Cardiovascular Health Study: This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, U01HL130114, R01HL087652, R01HL105756, R01HL103612,and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629, R01AG033193, R01AG15928, R01AG20098, and U01AG049505 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Framingham Heart Study. This work was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contracts N01-HC-25195 and HHSN268201500001I). This study was also supported by grants from the National Institute on Aging: R01AG033193, U01AG049505, U01AG52409, R01AG054076, RF1AG0059421 (S. Seshadri). S. Seshadri and A.L.D. were also supported by additional grants from the National Institute on Aging (R01AG049607, R01AG033040, RF1AG0061872, U01AG058589) and the National Institute of Neurological Disorders and Stroke (R01-NS017950, NS100605). The content is solely the responsibility of

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

the authors and does not necessarily represent the official views of the US National Institutes of Health.

QTLs/TWAS analyses. The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.synapse.org/). For MayoRNAseq, the study data were provided by the following sources: The Mayo Clinic Alzheimers Disease Genetic Studies, led by Dr. Nilufer Ertekin-Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimers Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinsons Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimers Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimers Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinsons Research. For ROSMAP, the study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG30146, R01AG36042 (5hC methylation, ATACseq), RC2AG036547 (H3K9Ac), R01AG36836 (RNAseq), R01AG48015 (monocyte RNAseq) RF1AG57473 (single nucleus RNAseq), U01AG32984 (genomic and whole exome sequencing), U01AG46152 (ROSMAP AMP-AD, targeted proteomics), U01AG46161(TMT proteomics), U01AG61356 (whole genome sequencing, targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at www.radc.rush.edu. For MSBB, the data were generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1: Summary of association results in Stage I, Stage II and Stage I + II for known loci with a genome-wide significant signal

Table 2: Summary of association results in Stage I, Stage II and Stage I + II for novel loci with a genome-wide significant signal

Table 3: Results for the most 20 significant pathways from MAGMA pathway analysis based on Stage I results

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1: Manhattan plot of the Stage I results. Variants in new loci are highlighted in red. Loci with a genome-wide significant signal are annotated (known loci in black and new loci in red). Variants with P value below $1x10^{-95}$ are not shown. The red dotted line represents the genome-wide significance level (P value = $5x10^{-8}$), while the black dotted line represents the suggestive significance level (P value = $1x10^{-5}$).

Figure 2: Top candidate genes in the 31 new genome-wide-significant loci (left panel) and level of expression in different brain cell types (right panel). In order to prioritize candidate genes in the new loci, we considered the "nearest" protein-coding gene from the lead variant and the genes exhibiting AD-related modulations within a region of 1 Mb around the lead variant. The average expression of each gene expressed by at least 10% of cells (pct.exp > 0.1) was rescaled from 0 to 2, allowing the identification of genes expressed by unique or multiple cell classes. Brown squares show significant hit for the respective column, dark green squares indicate the most probable candidate genes, bright greens squares indicate if several genes were retained in the same locus, and silver square indicate a prioritized gene in an independent known locus. The columns demonstrating lead variant m/haQTL effects in DLPFC and lead variant eQTL effects in monocyte and macrophages are annotated for the type of association (mQTL: methylation QTL, haQTL: histone acetylation QTL, mon: monocyte eQTL, mac: macrophage eQTL), and the superscript indicates that the associated feature is annotated for both genes. Arrow indicates level of expression of the candidate genes (full line for the most probable genes and dotted lines are for genes with less strong evidence).

Figure 3: QTL mapping, QTL colocalization, and TWAS results on SEC61G and TSPAN14 loci. (a) An intergenic, distant, and low-frequency cis-eQTL for *EGFR* locus colocalizes (PP4: 0.98) with EADB GWAS signal where genetic downregulation of *EGFR* is significantly associated with lower AD risk. (b) Elucidation of *TSPAN14* locus: through e/sQTLs, the protective signal is associated with decreased *TSPAN14* expression and increased cryptic splicing (within ADAM10-interacting domain) that are confirmed by long-read single-molecule sequencing on cDNA derived from hippocampus (Hipp.), frontal cortex (FC), lymphoblastoid cell lines (LCL). GWAS and e/sQTL association signals are shown as $-\log_{10}(P)$ and cumulative coverage tracks are in $\log_{10}$ scale. The purple dots show the lead variants in the investigated loci, and LD r^2 values were calculated with respect to these lead variants using the 1000 Genomes non-Finnish European LD reference panel prepared for the TWAS analyses.

Figure 4: Association of PRS with the risk to convert from MCI to AD or dementia in 10 cohorts and meta-analysis. PRS was based on the genetic data of 60 variants (see Methods and Supplementary Table 22).

Table 1

							Stage I			Stage II			Stage I + II				
Variant[a]	Chr.[b]	Position[c]	Gene[d]	Known locus	Minor/Major allele	MAF[e]	App. OR[f]	95% CI[g]	P[h]	OR[i]	95% CI[g]	P[h]	App. OR[f]	95% CI[g]	P[h]	I^2 (%)[j]	P[k]
rs4233366	1	161 189 357	ADAMTS4	ADAMTS4	T/C	0,263	0.95	0.93-0.97	$1.3x10^{-8}$	0.94	0.91-0.98	$1.2x10^{-3}$	0.95	0.93-0.96	$5.9x10^{-11}$	0	$9.1x10^{-1}$
rs679515	1	207 577 223	CR1	CR1	T/C	0,188	1.13	1.11-1.16	$5.2x10^{-33}$	1.14	1.10-1.19	$2.6x10^{-12}$	1.14	1.11-1.16	$1.2x10^{-43}$	21.5	$2.2x10^{-1}$
rs6733839	2	127 135 234	BIN1	BIN1	T/C	0,389	1.18	1.16-1.20	$6.5x10^{-90}$	1.19	1.16-1.23	$6.1x10^{-28}$	1.19	1.17-1.20	$5.6x10^{-116}$	32.6	$1.1x10^{-1}$
rs10933431	2	233 117 202	INPP5D	INPP5D	G/C	0,234	0.92	0.90-0.94	$1.0x10^{-17}$	0.94	0.91-0.98	$2.2x10^{-3}$	0.92	0.91-0.94	$1.7x10^{-19}$	0	$6.8x10^{-1}$
rs6846529	4	11 023 507	CLNK	CLNK/HS3ST1	C/T	0,283	1.07	1.05-1.09	$1.3x10^{-13}$	1.07	1.04-1.11	$7.1x10^{-5}$	1.07	1.05-1.09	$4.1x10^{-17}$	42.4	$4.7x10^{-2}$
rs6605556	6	32 615 322	HLA-DQA1	HLA	G/A	0,161	0.91	0.89-0.93	$1.0x10^{-17}$	0.90	0.85-0.95	$1.5x10^{-4}$	0.91	0.89-0.92	$7.7x10^{-21}$	0	$7.8x10^{-1}$
rs10947943	6	41 036 354	UNC5CL	TREM2	A/G	0,142	0.95	0.93-0.97	$6.2x10^{-6}$	0.92	0.88-0.96	$1.4x10^{-4}$	0.94	0.92-0.96	$8.1x10^{-9}$	9.8	$3.5x10^{-1}$
rs143332484	6	41 161 469	TREM2	TREM2	T/C	0,013	1.40	1.30-1.50	$6.0x10^{-19}$	1.72	1.45-2.03	$3.9x10^{-10}$	1.44	1.35-1.55	$1.7x10^{-26}$	53.7	$1.1x10^{-2}$
rs75932628	6	41 161 514	TREM2	TREM2	T/C	0,003	2.42	2.06-2.84	$1.4x10^{-27}$	2.30	1.80-2.94	$2.3x10^{-11}$	2.39	2.09-2.73	$2.5x10^{-37}$	0	$6.9x10^{-1}$
rs60755019	6	41 181 270	TREML2	TREM2	G/A	0,004	1.55	1.32-1.83	$1.8x10^{-7}$	1.51	1.02-2.25	$4.1x10^{-2}$	1.55	1.33-1.80	$2.1x10^{-8}$	49.9	$7.6x10^{-2}$
rs1385742	6	47 627 419	CD2AP	CD2AP	A/T	0,357	1.06	1.05-1.08	$4.7x10^{-13}$	1.09	1.06-1.13	$1.9x10^{-8}$	1.07	1.06-1.09	$1.7x10^{-19}$	0	$8.4x10^{-1}$
rs6976216	7	37 854 896	NME8	NME8	A/G	0,309	0.96	0.94-0.97	$2.3x10^{-7}$	0.94	0.91-0.97	$1.4x10^{-4}$	0.95	0.94-0.97	$2.2x10^{-10}$	0	$8.9x10^{-1}$
rs7384878	7	100 334 426	SPDYE3	ZCWPW1/NYAP1	C/T	0,310	0.93	0.91-0.94	$2.1x10^{-18}$	0.92	0.89-0.95	$1.0x10^{-6}$	0.92	0.91-0.94	$1.3x10^{-23}$	0	$5.8x10^{-1}$
rs10265814	7	143 424 528	EPHA1	EPHA1	C/G	0,499	1.06	1.04-1.07	$8.9x10^{-12}$	1.05	1.02-1.08	$1.7x10^{-3}$	1.06	1.04-1.07	$6.5x10^{-14}$	49.8	$2.1x10^{-2}$
rs73223431	8	27 362 470	PTK2B	PTK2B	T/C	0,369	1.07	1.05-1.09	$5.3x10^{-15}$	1.08	1.04-1.11	$3.7x10^{-6}$	1.07	1.05-1.09	$1.2x10^{-19}$	0	$5.9x10^{-1}$
rs867230	8	27 610 986	CLU	CLU	C/A	0,397	0.90	0.89-0.92	$1.5x10^{-33}$	0.92	0.89-0.95	$1.3x10^{-7}$	0.91	0.89-0.92	$1.7x10^{-39}$	18.2	$2.6x10^{-1}$
rs7912495	10	11 676 714	USP6NL	ECHDC3	G/A	0,462	1.06	1.04-1.08	$2.9x10^{-12}$	1.07	1.04-1.11	$8.2x10^{-6}$	1.06	1.05-1.08	$1.5x10^{-16}$	19.6	$2.4x10^{-1}$
rs10437655	11	47 370 397	SPI1	CELF1	A/G	0,399	1.06	1.04-1.08	$8.2x10^{-12}$	1.05	1.02-1.08	$1.5x10^{-3}$	1.06	1.04-1.07	$5.3x10^{-14}$	12.9	$3.1x10^{-1}$
rs1582763	11	60 254 475	MS4A4A	MS4A4A	A/G	0,371	0.92	0.90-0.93	$1.7x10^{-24}$	0.88	0.86-0.91	$1.2x10^{-14}$	0.91	0.90-0.92	$1.4x10^{-36}$	37.1	$8.0x10^{-2}$
rs10792832	11	86 156 833	PICALM	PICALM	A/G	0,358	0.90	0.89-0.91	$6.3x10^{-36}$	0.89	0.86-0.92	$1.5x10^{-13}$	0.90	0.88-0.91	$1.1x10^{-47}$	19.8	$2.4x10^{-1}$
rs11218343	11	121 564 878	SORL1	SORL1	C/T	0,039	0.85	0.81-0.88	$1.0x10^{-14}$	0.79	0.73-0.86	$6.5x10^{-8}$	0.84	0.81-0.87	$1.0x10^{-20}$	5.2	$3.9x10^{-1}$
rs17125924	14	52 924 962	FERMT2	FERMT2	G/A	0,089	1.09	1.06-1.12	$5.8x10^{-10}$	1.09	1.04-1.15	$9.7x10^{-4}$	1.09	1.07-1.12	$2.2x10^{-12}$	0	$9.6x10^{-1}$
rs12590654	14	92 472 511	SLC24A4	SLC24A4/RIN3	A/G	0,328	0.93	0.92-0.95	$2.1x10^{-15}$	0.92	0.89-0.95	$1.1x10^{-6}$	0.93	0.92-0.94	$1.5x10^{-20}$	0	$8.2x10^{-1}$
rs8032980	15	50 711 649	SPPL2A	SPPL2A	C/G	0,345	0.96	0.95-0.98	$4.7x10^{-6}$	0.94	0.91-0.97	$3.6x10^{-4}$	0.96	0.94-0.97	$1.1x10^{-8}$	19.1	$2.5x10^{-1}$
rs593742	15	58 753 575	ADAM10	ADAM10	G/A	0,295	0.94	0.92-0.96	$1.0x10^{-11}$	0.94	0.91-0.97	$4.0x10^{-4}$	0.94	0.93-0.96	$1.8x10^{-14}$	0	$6.6x10^{-1}$
rs117618017	15	63 277 703	APH1B	APH1B	T/C	0,144	1.12	1.09-1.15	$1.7x10^{-21}$	1.11	1.06-1.17	$4.9x10^{-5}$	1.12	1.09-1.14	$4.3x10^{-25}$	0	$6.4x10^{-1}$
rs889555	16	31 111 250	BCKDK	KAT8	T/C	0,281	0.95	0.93-0.96	$1.0x10^{-9}$	0.96	0.93-1.00	$3.1x10^{-2}$	0.95	0.93-0.96	$1.4x10^{-10}$	6.9	$3.8x10^{-1}$
rs4985556	16	70 660 097	IL34	IL34	A/C	0,115	1.06	1.03-1.09	$5.6x10^{-6}$	1.13	1.07-1.18	$4.1x10^{-6}$	1.07	1.05-1.10	$9.4x10^{-10}$	25.9	$1.8x10^{-1}$
rs12446759	16	81 739 398	PLCG2	PLCG2	G/A	0,403	0.94	0.93-0.96	$3.6x10^{-12}$	0.94	0.91-0.97	$3.1x10^{-4}$	0.94	0.93-0.96	$4.9x10^{-15}$	17.8	$2.6x10^{-1}$
rs72824905	16	81 908 423	PLCG2	PLCG2	G/C	0,008	0.74	0.67-0.81	$2.7x10^{-10}$	0.80	0.64-0.99	$3.7x10^{-2}$	0.75	0.68-0.81	$3.6x10^{-11}$	0	$5.4x10^{-1}$
rs7225151	17	5 233 752	SCIMP	SCIMP/RABEP1	A/G	0,124	1.09	1.06-1.12	$2.6x10^{-12}$	1.08	1.03-1.13	$7.0x10^{-4}$	1.09	1.07-1.11	$7.8x10^{-15}$	0	$5.2x10^{-1}$
rs199515	17	46 779 275	WNT3	MAPT	G/C	0,219	0.94	0.93-0.96	$6.0x10^{-9}$	0.93	0.90-0.97	$3.7x10^{-4}$	0.94	0.93-0.96	$1.0x10^{-11}$	62.3	$1.5x10^{-3}$
rs616338	17	49 219 935	ABI3	ABI3	T/C	0,012	1.34	1.24-1.45	$5.4x10^{-13}$	1.27	1.00-1.60	$4.8x10^{-2}$	1.33	1.24-1.44	$8.3x10^{-14}$	30.5	$1.4x10^{-1}$
rs2526377	17	58 332 680	TSPOAP1	TSPOAP1	G/A	0,445	0.96	0.94-0.97	$4.1x10^{-8}$	0.91	0.89-0.94	$1.3x10^{-8}$	0.95	0.93-0.96	$6.6x10^{-14}$	45.9	$3.1x10^{-2}$
rs4277405	17	63 471 557	ACE	ACE	C/T	0,384	0.93	0.92-0.95	$7.2x10^{-16}$	0.95	0.92-0.98	$6.5x10^{-4}$	0.94	0.92-0.95	$2.5x10^{-18}$	0	$8.6x10^{-1}$
rs12151021	19	1 050 875	ABCA7	ABCA7	A/G	0,336	1.11	1.09-1.13	$4.1x10^{-30}$	1.09	1.05-1.13	$1.3x10^{-6}$	1.11	1.09-1.13	$4.2x10^{-35}$	47.7	$2.8x10^{-2}$
rs17462136	20	56 412 160	CASS4	CASS4	C/G	0,086	0.89	0.86-0.91	$5.8x10^{-16}$	0.89	0.84-0.94	$1.9x10^{-5}$	0.89	0.87-0.91	$5.5x10^{-20}$	11.0	$3.3x10^{-1}$
rs2830489	21	26 775 872	ADAMTS1	ADAMTS1	T/C	0,281	0.95	0.93-0.96	$1.7x10^{-9}$	0.94	0.91-0.97	$4.5x10^{-4}$	0.95	0.93-0.96	$3.4x10^{-12}$	0	$5.6x10^{-1}$

[a]Reference SNP (rs) number according to dbSNP build 153; [b]Chromosome; [c]GRCh38 assembly; [d]Nearest protein coding gene according to Gencode release 33; [e]Weighted average of minor allele frequency across discovery studies; [f]Approximate odds-ratio calculated with respect to the minor allele; [g]95% confidence interval; [h]P value; [i]Odds-ratio calculated with respect to the minor allele; [j]Heterogeneity statistic; [k]P value for heterogeneity statistic

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2

							Stage I			Stage II			Stage I + II				
Locus number	Variant[a]	Chr.[b]	Position[c]	Gene[d]	Major/Minor allele	MAF[e]	App. OR[f]	95% CI[g]	P[h]	OR[i]	95% CI[g]	P[h]	App. OR[f]	95% CI[g]	P[h]	I^2 (%)[j]	P[k]
1	rs143080277	2	105 749 599	NCK2	C/T	0.005	1.48	1.31-1.66	$7.9x10^{-11}$	1.42	1.13-1.78	$2.3x10^{-3}$	1.47	1.32-1.63	$7.1x10^{-13}$	20	$2.7x10^{-1}$
2	rs61762319	3	155 084 189	MME	G/A	0.026	1.15	1.10-1.21	$2.1x10^{-8}$	1.30	1.15-1.47	$2.6x10^{-5}$	1.17	1.12-1.23	$1.3x10^{-11}$	5.5	$3.9x10^{-1}$
3	rs3822030	4	993 555	IDUA	G/T	0.429	0.95	0.93-0.97	$5.0x10^{-10}$	0.96	0.93-0.99	$4.3x10^{-3}$	0.95	0.94-0.96	$8.3x10^{-12}$	2.7	$4.2x10^{-1}$
4	rs250435	5	14 664 712	OTULIN	G/A	0.079	1.08	1.05-1.12	$4.2x10^{-7}$	1.13	1.07-1.21	$6.2x10^{-5}$	1.09	1.06-1.13	$2.5x10^{-10}$	35.9	$8.8x10^{-1}$
5	rs62375397	5	87 002 714	RASA1	T/C	0.210	1.08	1.06-1.10	$8.6x10^{-14}$	1.06	1.02-1.10	$1.3x10^{-3}$	1.07	1.06-1.09	$5.2x10^{-16}$	45.3	$3.3x10^{-1}$
6	rs113706587	5	180 201 150	RASGEF1C	A/G	0.110	1.10	1.07-1.13	$3.4x10^{-12}$	1.10	1.05-1.16	$1.9x10^{-4}$	1.10	1.07-1.12	$2.8x10^{-15}$	0	$8.4x10^{-1}$
7	rs785129	6	114 291 731	HS3ST5	T/C	0.350	1.05	1.03-1.06	$2.9x10^{-7}$	1.04	1.00-1.07	$3.1x10^{-2}$	1.04	1.03-1.06	$3.0x10^{-8}$	10.6	$3.4x10^{-1}$
8	rs6943429	7	7 817 263	UMAD1	T/C	0.421	1.04	1.03-1.06	$3.0x10^{-7}$	1.06	1.02-1.09	$6.2x10^{-4}$	1.05	1.03-1.06	$9.0x10^{-10}$	50.3	$1.6x10^{-1}$
9	rs10952097	7	8 204 382	ICA1	T/C	0.114	1.07	1.04-1.10	$3.3x10^{-7}$	1.08	1.03-1.14	$3.1x10^{-3}$	1.08	1.05-1.10	$3.8x10^{-9}$	52.2	$1.4x10^{-1}$
10	rs7781670	7	12 211 934	TMEM106B	G/C	0.409	0.96	0.94-0.98	$6.1x10^{-7}$	0.96	0.93-0.99	$8.6x10^{-3}$	0.96	0.95-0.97	$1.7x10^{-8}$	42.8	$4.5x10^{-1}$
11	rs76928645	7	54 873 635	SEC61G	T/C	0.103	0.93	0.90-0.95	$1.5x10^{-8}$	0.95	0.90-1.00	$4.9x10^{-2}$	0.93	0.91-0.95	$2.9x10^{-9}$	21.6	$2.2x10^{-1}$
12	rs4731	8	11 808 828	FDFT1	G/A	0.081	1.08	1.05-1.11	$7.6x10^{-8}$	1.06	1.00-1.12	$4.3x10^{-2}$	1.08	1.05-1.11	$1.2x10^{-8}$	54.4	$9.7x10^{-1}$
13	rs13276936	8	94 983 473	NDUFAF6	T/C	0.466	1.05	1.03-1.07	$2.8x10^{-9}$	1.04	1.01-1.07	$1.0x10^{-2}$	1.05	1.03-1.06	$1.1x10^{-10}$	0	$5.1x10^{-1}$
14	rs34173062	8	144 103 704	SHARPIN	A/G	0.081	1.12	1.09-1.16	$2.9x10^{-12}$	1.17	1.08-1.27	$1.8x10^{-4}$	1.13	1.09-1.16	$3.7x10^{-15}$	44.6	$4.2x10^{-1}$
15	rs1800978	9	104 903 697	ABCA1	G/C	0.130	1.07	1.05-1.10	$1.6x10^{-8}$	1.05	1.00-1.10	$3.7x10^{-2}$	1.07	1.04-1.09	$2.5x10^{-9}$	27.7	$1.6x10^{-1}$
16	rs7068231	10	60 025 170	ANK3	T/G	0.403	0.95	0.94-0.97	$6.8x10^{-9}$	0.96	0.93-0.99	$1.0x10^{-2}$	0.95	0.94-0.97	$2.5x10^{-10}$	0	$5.9x10^{-1}$
17	rs6586028	10	80 494 228	TSPAN14	C/T	0.196	0.92	0.91-0.94	$1.3x10^{-14}$	0.94	0.91-0.98	$2.3x10^{-3}$	0.93	0.91-0.94	$1.7x10^{-16}$	0	$4.8x10^{-1}$
18	rs6584063	10	96 266 650	BLNK	G/A	0.046	0.89	0.86-0.93	$1.8x10^{-7}$	0.85	0.78-0.93	$2.6x10^{-4}$	0.89	0.85-0.92	$3.4x10^{-10}$	0	$9.3x10^{-1}$
19	rs3919533	12	7 055 497	C1S	C/T	0.172	0.95	0.93-0.97	$1.0x10^{-6}$	0.94	0.91-0.98	$5.7x10^{-3}$	0.95	0.93-0.97	$1.9x10^{-8}$	28.5	$1.6x10^{-1}$
20	rs74091721	14	106 638 262	IGH gene	T/G	0.145	0.94	0.92-0.96	$6.7x10^{-7}$	0.94	0.89-0.98	$3.1x10^{-3}$	0.94	0.92-0.96	$7.7x10^{-9}$	0	$9.9x10^{-1}$
21	rs2009833	15	49 901 356	ATP8B4	A/G	0.368	0.95	0.94-0.97	$3.1x10^{-8}$	0.96	0.93-0.99	$5.9x10^{-3}$	0.95	0.94-0.97	$6.5x10^{-10}$	0	$6.5x10^{-1}$
22	rs3848143	15	64 131 307	SNX1	G/A	0.220	1.05	1.03-1.07	$1.1x10^{-6}$	1.07	1.03-1.11	$3.0x10^{-4}$	1.05	1.04-1.07	$2.0x10^{-9}$	0	$5.5x10^{-1}$
23	rs12592898	15	78 936 857	CTSH	A/G	0.133	0.94	0.92-0.97	$1.3x10^{-6}$	0.93	0.89-0.98	$4.5x10^{-3}$	0.94	0.92-0.96	$2.1x10^{-8}$	3.7	$4.1x10^{-1}$
24	rs9783783	16	30 067 171	ALDOA	C/G	0.492	0.96	0.94-0.97	$4.5x10^{-8}$	0.97	0.94-1.00	$4.8x10^{-2}$	0.96	0.95-0.97	$8.1x10^{-9}$	0	$8.9x10^{-1}$
25	rs450674	16	79 574 511	MAF	C/T	0.373	0.96	0.94-0.97	$1.1x10^{-7}$	0.97	0.94-1.00	$3.5x10^{-2}$	0.96	0.94-0.97	$1.3x10^{-8}$	14.1	$3.0x10^{-1}$
26	rs56407236	16	90 103 687	PRDM7	A/G	0.069	1.12	1.08-1.15	$1.3x10^{-11}$	1.10	1.02-1.17	$9.0x10^{-3}$	1.11	1.08-1.14	$4.4x10^{-13}$	0	$6.0x10^{-1}$
27	rs5848	17	44 352 876	GRN	T/C	0.289	1.07	1.05-1.09	$1.8x10^{-12}$	1.04	1.01-1.08	$1.2x10^{-2}$	1.06	1.05-1.08	$1.2x10^{-13}$	0	$9.3x10^{-1}$
28	rs8103552	19	49 955 674	SIGLEC11	A/C	0.240	1.05	1.03-1.07	$5.7x10^{-6}$	1.09	1.05-1.13	$1.2x10^{-6}$	1.06	1.04-1.08	$2.5x10^{-10}$	0	$4.8x10^{-1}$
29	rs12984029	19	54 265 512	LILRB5	A/G	0.235	1.06	1.04-1.09	$3.2x10^{-9}$	1.05	1.01-1.09	$9.8x10^{-3}$	1.06	1.04-1.08	$1.2x10^{-10}$	33	$1.2x10^{-1}$
30	rs6742	20	63 743 088	SLC2A4RG	T/C	0.221	0.95	0.93-0.97	$2.5x10^{-6}$	0.93	0.89-0.98	$2.1x10^{-3}$	0.95	0.93-0.97	$2.6x10^{-8}$	0	$9.6x10^{-1}$
31	rs2154481	21	26 101 558	APP	C/T	0.476	0.95	0.94-0.97	$1.0x10^{-9}$	0.96	0.93-0.99	$7.6x10^{-3}$	0.95	0.94-0.97	$3.0x10^{-11}$	5.1	$4.0x10^{-1}$

[a]Reference SNP (rs) number according to dbSNP build 153; [b]Chromosome; cGRCh38 assembly; [d]Nearest protein coding gene according to Gencode release 33; [e]Weighted average of minor allele frequency across discovery studies; [f]Approximate odds-ratio calculated with respect to the minor allele; [g]95% confidence interval; [h]P value; [i]Odds-ratio calculated with respect to the minor allele; [j]Heterogeneity statistic; [k]P value for heterogeneity statistic

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

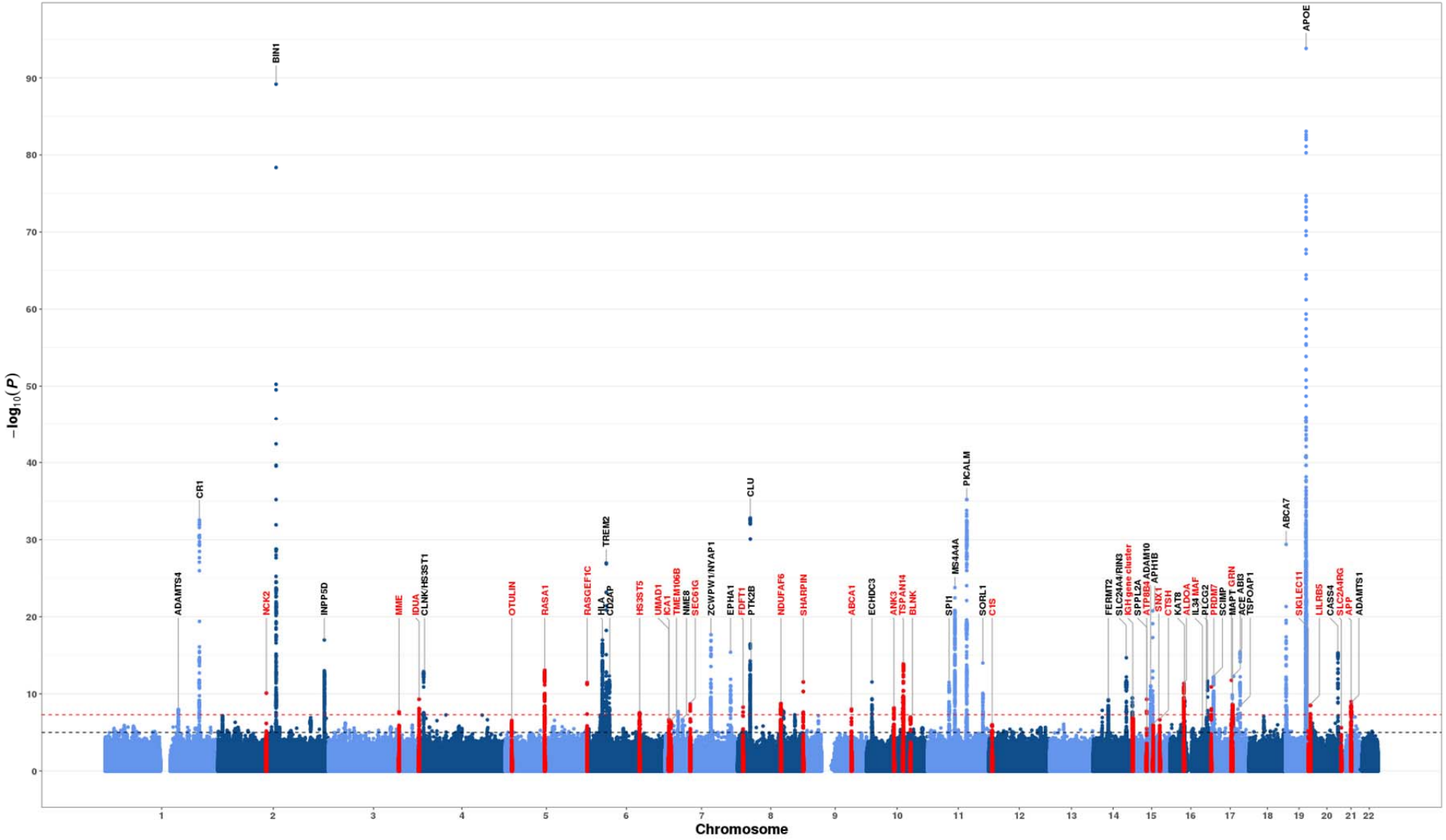
Table 3

Pathway	Number of genes in the pathway in the dataset	P value	q value[a]	Pathway description
GO:1902992	14	$9.5x10^{-17}$	$7.5x10^{-13}$	Negative regulation of amyloid precursor protein catabolic process
GO:1902430	12	$1.8x10^{-12}$	$7.3x10^{-9}$	Negative regulation of amyloid-beta formation
GO:0048156	12	$3.6x10^{-12}$	$9.4x10^{-9}$	Tau protein binding
R-HSA-5619055	1	$8.3x10^{-11}$	$1.6x10^{-7}$	Defective SLC24A4 causes hypomineralized amelogenesis imperfecta (AI)
GO:1905245	11	$9.8x10^{-10}$	$1.6x10^{-6}$	Regulation of aspartic-type peptidase activity
GO:1902991	32	$1.6x10^{-9}$	$2.1x10^{-6}$	Regulation of amyloid precursor protein catabolic process
GO:1902003	27	$5.0x10^{-9}$	$5.7x10^{-6}$	Regulation of amyloid-beta formation
GO:0032994	26	$1.0x10^{-8}$	$1.0x10^{-5}$	Protein-lipid complex
GO:0033700	12	$3.7x10^{-7}$	$2.9x10^{-4}$	Phospholipid efflux
GO:0034358	23	$4.1x10^{-7}$	$2.9x10^{-4}$	Plasma lipoprotein particle
GO:1990777	23	$4.1x10^{-7}$	$2.9x10^{-4}$	Lipoprotein particle
R-HSA-198933	121	$8.2x10^{-7}$	$5.4x10^{-4}$	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell
GO:0034364	18	$1.2x10^{-6}$	$7.1x10^{-4}$	High-density lipoprotein particle
GO:0042116	28	$2.0x10^{-6}$	$1.2x10^{-3}$	Macrophage activation
GO:0071825	41	$2.8x10^{-6}$	$1.5x10^{-3}$	Protein-lipid complex subunit organization
GO:0001774	18	$3.2x10^{-6}$	$1.5x10^{-3}$	Microglial cell activation
GO:0002269	18	$3.2x10^{-6}$	$1.5x10^{-3}$	Leukocyte activation involved in inflammatory response
GO:0033003	24	$3.8x10^{-6}$	$1.7x10^{-3}$	Regulation of mast cell activation
GO:0045332	22	$4.3x10^{-6}$	$1.7x10^{-3}$	Phospholipid translocation
GO:0044433	1132	$4.4x10^{-6}$	$1.7x10^{-3}$	Cytoplasmic vesicle part

[a]significant after FDR correction (q value ≤ 0.05)

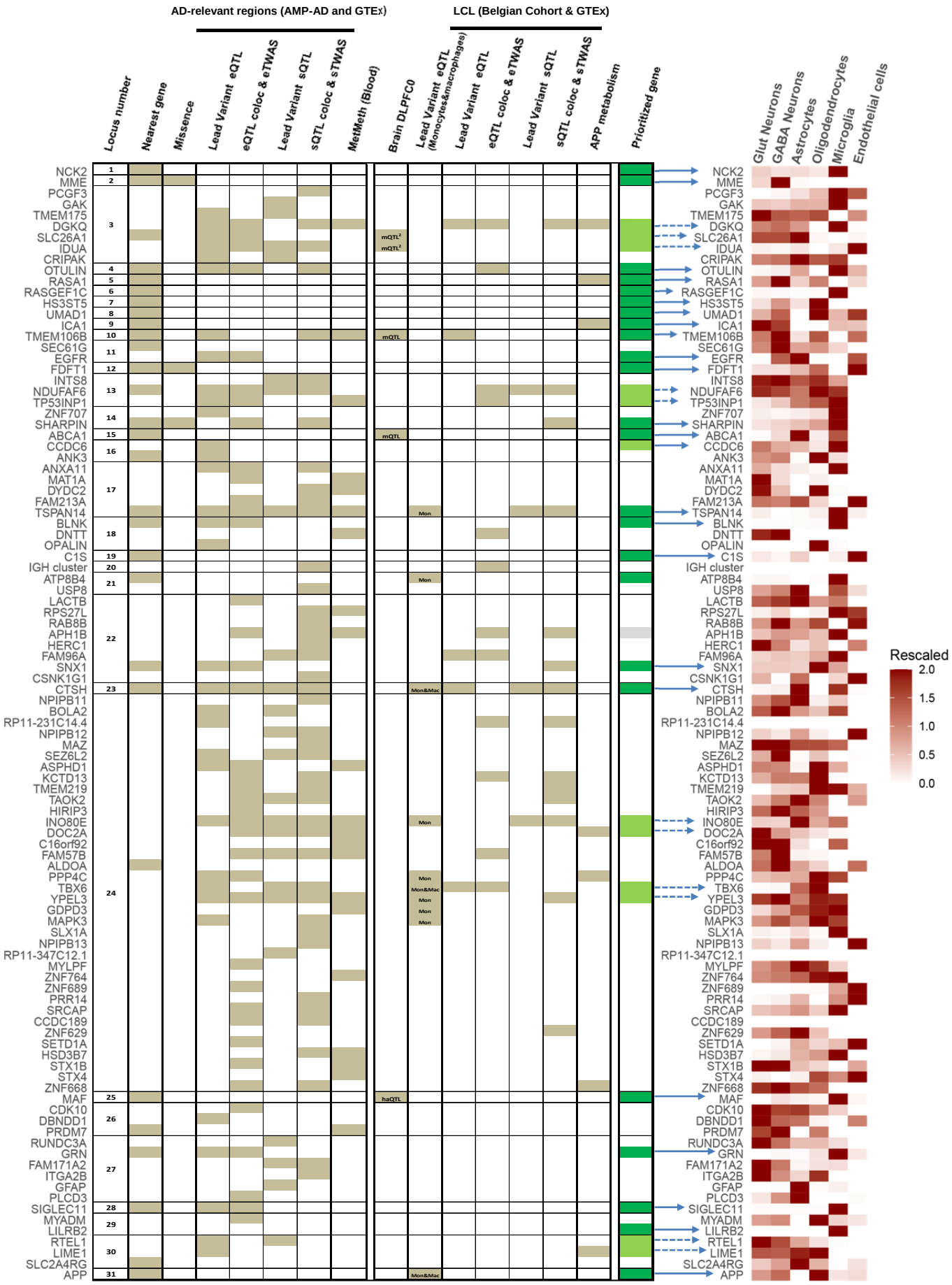
medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1



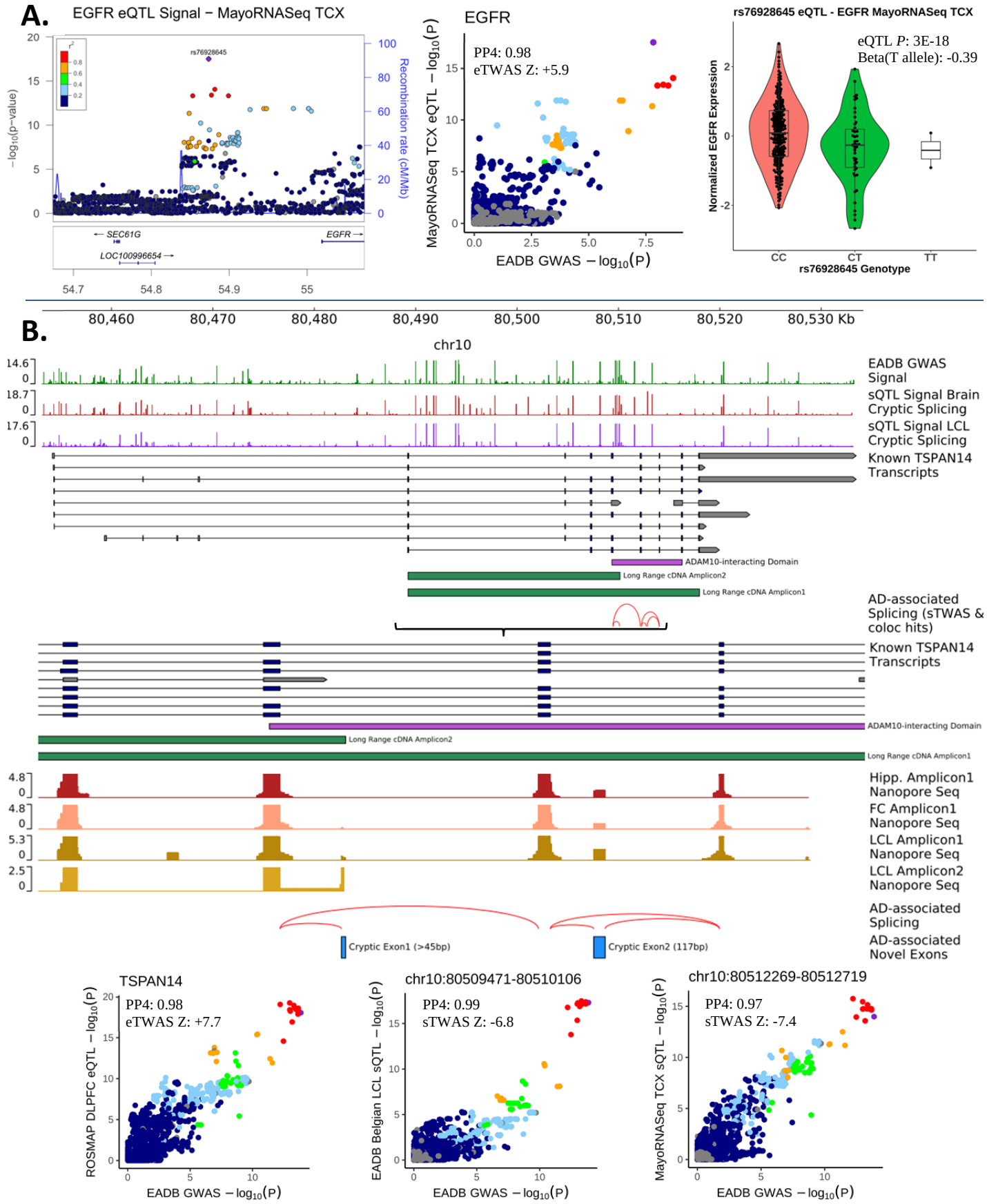
medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2



medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3 :



medRxiv preprint doi: https://doi.org/10.1101/2020.10.01.20200659; this version posted October 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

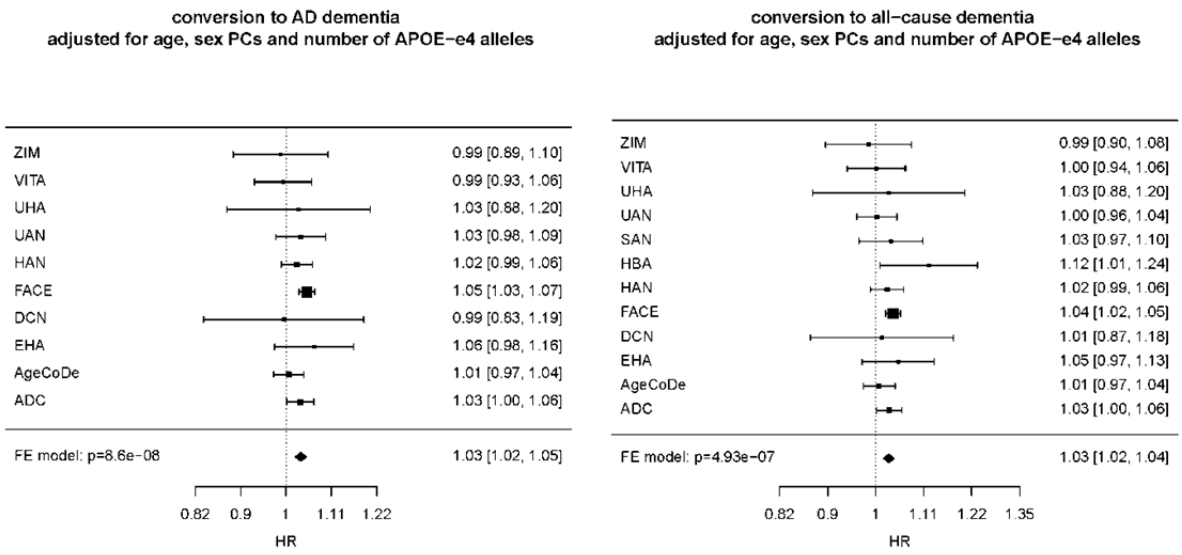



Figure 4