Title
Genetic Risk Factors of Tauopathies

Author
Chen, Jason Andrew

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Genetic Risk Factors of Tauopathies

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy in Bioinformatics

by

Jason Andrew Chen

2017

ABSTRACT OF THE DISSERTATION

Genetic Risk Factors of Tauopathies

by

Jason Andrew Chen

Doctor of Philosophy in Bioinformatics

University of California, Los Angeles, 2017

Professor Giovanni Coppola, Chair

The tauopathies are a group of neurodegenerative diseases characterized by abnormal tau neuropathology and include frontotemporal dementia (FTD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), and Alzheimer's disease (AD), collectively with a heavy burden on the lives of patients and on public health. Uncertainty of the underlying disease mechanisms hinders the development of effective treatments and diagnostics. While most cases of each disease are sporadic, a strong genetic contribution has been identified and provides an opening for the discovery of clues to their pathophysiology.

I led a series of genetic screens to systematically search for genetic contributors to the tauopathies. A genome-wide association study of PSP was performed to pinpoint common polymorphisms that contribute to disease risk. Aside from supporting associations at loci that had already been reported near *MAPT*, *STX6*, *MOBP*, and *EIF2AK3*, novel associations were discovered at genome-wide significance (near *RUNX2* and *SLCO1A2*) and suggestive significance (near *DUSP10*, *SP1*, *ASAP1*, and *WDR63*). Furthermore, we identified genetic correlations between PSP, Parkinson's disease, and amyotrophic lateral sclerosis, indicating that the role of tau may extend to more neurodegenerative diseases than currently appreciated.

Leveraging the coverage of the genotyping array, copy number variants were also called in a subset of this disease cohort. A previously unreported duplication spanning part of the tau gene was found in two PSP patients, expanding the scope of mutations found in the disease.

While the SNP genotyping array in the previous study provided genome-wide coverage of common polymorphisms, a newly designed exome array specifically typed rare exonic variants found in large population cohorts. We performed one of the first exome array studies aimed at uncovering causal genetic risk factors in AD, FTD, and PSP. We identified a contribution of exonic variants in the *ABCA7* gene to AD, presaging a body of current literature focusing on protein-disrupting mutations of this gene. Additionally, we provided support for the candidate genes *PAXIP1* and *DYSF* in AD. Whole genome sequencing studies outlined a contribution of rare protein-disrupting mutations in predicted damaging genes. The *MAPT* A152T variant was also confirmed as a risk factor in tauopathies, with our work expanding the phenotypic spectrum.

We further established the downstream molecular effects of tauopathy risk alleles. The H1 haplotype of chromosome 17q21.31, a major risk factor for PSP and CBD that is also polymorphic in the general population, had a large influence on DNA methylation in the region and mRNA expression of nearby genes. This methylation pattern appeared to mediate some of the risk conferred from the H1 haplotype. We also identified another methylation signature on the promoter of the gene encoding IL-1β that was correlated with aging and methylation, with mechanistic support from mouse models of sirtuin 1 function. Finally, we linked rs242557 - a risk allele within *MAPT* - with plasma tau concentration, a potential biomarker of clinical utility.

Future work will focus on improving statistical power by increasing the size of sample cohorts and integrating additional layers of genomic data, including those from public epigenomics and mRNA expression studies. I outline a path for the translation of the results from completed and future work to enabling precision medicine for patients suffering from tauopathies.

The dissertation of Jason Chen is approved.

Jorge Lazareff

Stefan Horvath

Daniel Geschwind

Giovanni Coppola, Committee Chair

University of California, Los Angeles

2017

To my parents, Gary and Sharon, and my brother William; and to all of our patients and their families, who have enabled and inspired this work.

Table of Contents

List of Figures & Tables

Acknowledgement

From my father, I learned the primacy of truth, the advantages of mathematical ability, and the joy of a hard day's work.

From my Ph.D. supervisor and committee chair, Prof. Giovanni Coppola, I was encouraged to explore the frontiers of science and saw what is possible with curiosity, willpower, and collaboration. He pushed me to dare greatly, gave me room to fail and to learn, and thrust opportunities at me from the moment I stepped foot in his lab.

From my co-supervisor, Prof. Dan Geschwind, I observed the dedication and wisdom needed to be a true leader and a visionary in the field. To push against the status quo and light my own path forward.

From the foremost pediatric neurosurgeon not (yet) in the Presidential order of succession, Prof. Jorge Lazareff, I learned to give hope to patients and their families, and to question orthodoxy at the origins of its ideas. He saw the intrinsic value in all of his children, from the M.D./Ph.D. student to the conjoined twin girls of Guatemala to abandoned infants in Shanghai with neural tube defects.

From Prof. Steve Horvath, a pioneer who laid the groundwork for some of the most important discoveries in genomics, to trek toward the hidden biology in data.

From my fellow trainees: Alden Huang, Grant Belgard, Neel Parikshak, Mochtar Pribadi, to understand genomics and keep abreast of the latest developments. Most of my knowledge and practical skills were gained not from books or journal articles, but from our conversations.

From Alice Zhang, my first friend in medical school and with who I started a company, to hustle and achieve my goals.

From my colleagues at Verge Genomics, Victor Hanson-Smith, David Lamparter, Sarah Silvergleid, Raj Bhatnagar, Nick Wisniewski, and Katja Hebestreit, to find passion in life and strive toward making the world a better place. It still boggles my mind that such unbelievably talented people would work at our fledgling company.

From my erstwhile muse, Annette, to love and to work as a team.

The kindness of the administrative staff and lab managers and their willingness to support me have always humbled me, and I would like to acknowledge the important contributions of Margaret Chu, Lauren Kawaguchi, Jenifer Sakai, Helena Tran, and Vanessa Marrero.

This work is the culmination of the efforts of a number of incredible co-authors and collaborators:

Chapter 1 consists of my thoughts on the importance of this work and a compilation of some of the literature I have reviewed in regards to the pathology and genetics of tauopathies. I put together the literature review with suggestions from Adam Boxer and Giovanni Coppola, and it is concurrently being prepared for publication. Outside of the literature review, some of the text for rationale and design of the study was included from a grant application that I authored,

ultimately awarded as NINDS Ruth L. Kirschstein National Research Service Award F31 NS084556.

Chapter 2 is adapted from two currently unpublished works, tentatively titled 'Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases' (manuscript in preparation) and 'Identification of *MAPT*, *MOBP* and *STX6* as susceptibility loci in progressive supranuclear palsy' (manuscript in submission). The first describes a meta-analysis of PSP genome-wide association studies that I performed. Some of the data used in the study was acquired from co-authors, Zhongbo Chen, Jennifer Lowe, Jennifer Yokoyama, Gilbert Bensimon, Nigel Leigh, Christine Payan, Ammar Al-Chalabi, and Adam Boxer. From our group, overseen by Daniel Geschwind and Giovanni Coppola, Hyejung Won helped to perform the Hi-C analysis, Alden Huang gave suggestions and guidance about the statistical aspects of the manuscript, and Kevin Wojta helped with plating of the samples and logistical coordination. The second manuscript describes a GWAS performed by Ammar al-Chalabi and his group, with co-authors Zhongbo Chen, Aleksey Shatunov, Ashley R. Jones, Cathryn M. Lewis, Christine A. M. Payan, Wolfgang Lieb, Andre Franke, Panagiotis Deloukas, Philippe Amouyel, Christophe Tzourio, Jean-François Dartigues, NNIPPS and BBBIPPS Study Groups, Albert Ludolph, Gilbert Bensimon, P. Nigel Leigh, and Jeff M. Bronstein. The section that was adapted covers work describing genome-wide copy-number variation analysis that I performed with the help of Jennifer Lowe (who organized and directed the logistics of the samples), Alden Huang (who helped in the statistical analysis), and Stephanie Kravitz (who performed the qPCR validation experiments), and was overseen by Giovanni Coppola and Daniel Geschwind.

Chapter 3 summarizes work described in a publication entitled 'A multiancestral genome-wide exome array study of Alzheimer disease, frontotemporal dementia, and progressive

supranuclear palsy' (Chen et al., 2015). Co-authors included members of our group, Qing Wang, Jeremy Davis-Turak, Yun Li, Sandy Chan Hsu, Renee Sears, Doxa Chatzopoulou, Alden Huang, Eric Klein, and Kevin Wojta, who maintained the GIFT collection and database for organizing these important samples; and collaborators who provided samples and performed clinical evaluations, including Anna Karydas, Jason Lee, Duane L. Beekly, Adam Boxer, Kelley M. Faber, Claudia M. Haase, Josh Miller, Wayne W. Poon, Ami Rosen, Howard Rosen, Anna Sapozhnikova, Jill Shapira, Arousiak Varpetian, Tatiana M. Foroud, Robert W Levenson, Allan I. Levey, Walter A. Kukull, Mario F. Mendez, John Ringman, Helena Chui, Carl Cotman, Charles DeCarli, and Bruce L. Miller. The work was supervised by Giovanni Coppola and Daniel Geschwind. Permission to reproduce the manuscript is granted from Wolters Kluwer.

Chapter 4 previews work from a long-running study of the genome sequencing of PSP, overseen by Daniel Geschwind and Giovanni Coppola. A brief section of Chapter 4 is adapted from a paper that I co-authored, entitled 'A152T tau allele causes neurodegeneration that can be ameliorated in a zebrafish model by autophagy induction' (Lopez et al., 2017). I performed statistical analysis of the prevalence of the *MAPT* A152T variant in various disease cohorts, the section that was included in this work. Co-authors of the work include Ana Lopez, Suzee E. Lee, Kevin Wojta, Eliana Marisa Ramos, Eric Klein, Adam L. Boxer, Maria Luisa Gorno-Tempini, Daniel H. Geschwind, Lars Schlotawa, Nikolay V. Ogryzko, Eileen H. Bigio, Emily Rogalski, Sandra Weintraub, Marsel M. Mesulam; the work was led by Angeleen Fleming, Giovanni Coppola, Bruce L. Miller, and David C. Rubinsztein. The original manuscript was published under a Creative Commons Attribution (CC BY) license, permitting its use here.

Chapter 5 describes work published in a pair of publications. The former, 'SIRT1 deficiency in microglia contributes to cognitive decline in aging and neurodegeneration via epigenetic regulation of IL-1β' (Cho et al., 2015), was written largely by first author Seo-Hyun Choi, and co-

authors Faten Sayed, Michael E Ward, Fuying Gao, Thi A Nguyen, Grietje Krabbe, Peter Dongmin Sohn, Iris Lo, Sakura Minami, Nino Devidze, and Yungui Zhou, and supervised by Li Gan. I had the opportunity to contribute a "buzzer beater" describing in humans the methylation patterns that they observed in mouse, which is the section that is adapted. Giovanni Coppola oversaw this portion of the work. Sections of the article are reprinted here with permission from the publisher. The latter, entitled 'An epigenetic signature in peripheral blood associated with the haplotype on 17q21. 31, a risk factor for neurodegenerative tauopathy' (Li et al., 2014), describes the methylation signature associated with PSP and the extended PSP risk haplotype on 17q21.31. Yun Li was a co-first author of this manuscript and initially described the finding that the risk haplotype was highly associated with methylation at particular loci. Co-authors Renee L. Sears, Fuying Gao, Eric D. Klein, Anna Karydas, Michael D. Geschwind, Howard J. Rosen, Adam L. Boxer, Weilong Guo, Matteo Pellegrini, Steve Horvath, Bruce L. Miller, and Daniel H. Geschwind performed sample collection and preparation. Giovanni Coppola directed the work. The original manuscript was published under a Creative Commons Attribution (CC BY) license, permitting its use here.

Chapter 6 is partially comprised of results from a manuscript entitled 'Genome-wide association study identifies MAPT locus influencing human plasma tau levels' in which I performed a genome-wide association study of plasma tau in the ADNI cohort (Chen et al., 2017). The initial finding and much of the analysis was performed by co-first author Jin-Tai Yu. Kevin Wojta from our group performed genotyping work on the replication cohort. Henrik Zetterberg and Kaj Blennow contributed the measurements of plasma tau concentrations and their long expertise in this field. Other co-authors who contributed to the analyses include Hui-Fu Wang, Jennifer S. Yokoyama, Michael W. Weiner, Joel H. Kramer, Howard Rosen, and Bruce L. Miller. The work was overseen by Adam Boxer and Giovanni Coppola. The manuscript is reproduced here with permission from the publisher.

Chapter 7 contains, amongst my hand-waving and pondering, work in submission led by Justin Ichida. I directed the analysis effort that is mentioned in the manuscript, with great contributions from my colleagues at Verge, Nick Wisniewski, Victor Hanson-Smith, T. Grant Belgard, Raj Bhatnagar, and Katja Hebestreit, and Alice Zhang who connected our groups and made this collaboration possible. It also includes unpublished work entitled 'Gene co-expression network analysis implicates microRNA processing in Parkinson's disease pathogenesis'.

My work has been generously funded by a number of groups. My stipend has been funded by taxpayers of the great United States of America and the State of California, through the National Institutes of Health grants: NIGMS Training Grant T32 GM08042, and NINDS Ruth L. Kirschstein National Research Service Award F31 NS084556 and through institutional support at UCLA; and by the Rainwater Charitable Foundation and the generosity of the Rainwater family.

Finally, I would like to dedicate this work to all of our patients and their families. I acknowledge "the patients" as a matter of habit in most of my publications, but their centrality to this effort cannot be overstated. It is the patients that inspire us to reveal secrets of the genome; it is the patients who have supported our research and believed in us when no one else would; and it is the patients who have donated their bodies and their data so that we may improve the lives of future generations.

"Among the Quadi, at the Granua"

References

Chen, J., Yu, J.-T., Wojta, K., Wang, H.-F., Zetterberg, H., Blennow, K., Yokoyama, J.S., Weiner, M.W., Kramer, J.H., Rosen, H.*, et al.* (2017). Genome-wide association study identifies MAPT locus influencing human plasma tau levels. Neurology *88*, 669-676.

Chen, J.A., Wang, Q., Davis-Turak, J., and et al. (2015). A multiancestral genome-wide exome array study of alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy. JAMA Neurology *72*, 414-422.

Cho, S.-H., Chen, J.A., Sayed, F., Ward, M.E., Gao, F., Nguyen, T.A., Krabbe, G., Sohn, P.D., Lo, I., Minami, S.*, et al.* (2015). SIRT1 Deficiency in Microglia Contributes to Cognitive Decline in Aging and Neurodegeneration via Epigenetic Regulation of IL-1β. The Journal of Neuroscience *35*, 807-818.

Li, Y., Chen, J.A., Sears, R.L., Gao, F., Klein, E.D., Karydas, A., Geschwind, M.D., Rosen, H.J., Boxer, A.L., Guo, W.*, et al.* (2014). An Epigenetic Signature in Peripheral Blood Associated with the Haplotype on 17q21.31, a Risk Factor for Neurodegenerative Tauopathy. PLoS Genetics *10*, e1004211.

Lopez, A., Lee, S.E., Wojta, K., Ramos, E.M., Klein, E., Chen, J., Boxer, A.L., Gorno-Tempini, M.L., Geschwind, D.H., Schlotawa, L.*, et al.* (2017). A152T tau allele causes neurodegeneration that can be ameliorated in a zebrafish model by autophagy induction. Brain *140*, 1128-1146.

VITA

2006 – 2010	B.S.E. Biomedical Engineering, A.B. Chemistry, Duke University, Durham, NC (*summa cum laude*) • Tau Beta Pi • Da Vinci Award in Biomedical Engineering
2010 – present	Medical Scientist Training Program, David Geffen School of Medicine at the University of California, Los Angeles
2012 – 2017	Graduate Student Researcher, Coppola Lab, University of California, Los Angeles • NIH NINDS Ruth L. Kirschstein National Research Service Award (F31 NS084556): Genetic Determinants of Tauopathies
2014 – 2017	Co-Founder & Chief Scientific Officer, Verge Genomics, San Francisco, CA • Y Combinator Summer 2015 Cohort

PUBLICATIONS

Chen J*, Yu J-T*, Wojta K, Wang H-F, Zetterberg H, Blennow K, Yokoyama JS, Weiner MW, Kramer JH, Rosen H, Miller BL, Coppola G, Boxer AL. (2017). Genome-wide association study identifies MAPT locus influencing human plasma tau levels. *Neurology* 88, 669-676.

Lopez A, Lee SE, Wojta K, Ramos EM, Klein E, Chen J, Boxer AL, Gorno-Tempini ML, Geschwind DH, Schlotawa L, Ogryzko NV, Bigio EH, Rogalski E, Weintraub S, Mesulam MM, Fleming A, Coppola G, Miller BL, Rubinsztein DC. (2017). A152T tau allele causes neurodegeneration that can be ameliorated in a zebrafish model by autophagy induction. *Brain* 140(4):1128-1146.

Kubota K, Chen J, Little M. (2016). Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. Movement *Disorders* 31(9):1314-1326.

Chen J, Penagarikano O, Belgard TG, Swarup V, Geschwind D. (2015). The Emerging Picture of Autism Spectrum Disorder: Genetics and Pathology. *Annual Review of Pathology: Mechanisms of Disease* 10(1):111-144.

Chen J, Wang Q, Davis-Turak J, Li Y, Karydas A, Hsu S, Sears R, Chatzopoulou D, Huang A, Wojta K, Klein E, Lee J, Beekly D, Boxer A, Faber K, Haase C, Miller J, Poon W, Rosen A, Rosen H, Sapozhnikova A, Shapira J, Varpetian A, Foroud T, Levenson R, Levey A, Kukull W, Mendez M, Ringman J, Chui H, Cotman C, DeCarli C, Miller B, Geschwind D, Coppola G.

(2015). A multi-ethnic genome-wide exome array study of Alzheimer's disease, frontotemporal dementia, and progressive supranuclear palsy. *JAMA Neurology* 72(4):414-422.

Chen J, Coppola G. (2015). Imaging Genetics. *Brain Mapping: An Encyclopedic Reference*. Toga, A (ed). Volume 3: 1037-1047. Academic Press: Elsevier.

Cho S-H, Chen J, Sayed F, Ward M, Gao F, Nguyen T, Krabbe G, Sohn P, Lo I, Minami S, Devidze N, Zhou Y, Coppola G, Gan L. (2015). SIRT1 deficiency in microglia contributes to cognitive decline in aging and neurodegeneration via epigenetic regulation of IL-1β. *The Journal of Neuroscience* 35(2):807-818.

Li Y*, Chen J*, Sears R, Gao F, Klein E, Karydas A, Geschwind M, Rosen H, Boxer A, Miller B, Geschwind D, Coppola G. (2014). An epigenetic signature in peripheral blood associated with neurodegenerative tauopathy and the risk-associated haplotype on 17q21.31. *PLOS Genetics* 10(3):e1004211.

Cheok S, Chen J, Lazareff J. (2014). The truth and coherence behind the concept of overdrainage of cerebrospinal fluid in hydrocephalic patients. *Child's Nervous System* 30(4):599-606.

Chen J, Lazareff J. (2014). Correction of Chiari malformation due to closure of a concomitant thoracic meningocele. *Child's Nervous System* 30(3):531-534.

Chen J, Coutin-Churchman P, Nuwer M, Lazareff J. (2012). Suboccipital craniotomy for Chiari I results in evoked potential conduction changes. *Surgical Neurology International* 3(1):165.

Chen J, Cho J, Choi B. (2007). A Convenient Truth: A model for sea level rise forecast. *The UMAP Journal* 29(3): 225.

PATENTS

Chen JA, Zhang AX. Motor-associated neurodegenerative disease and methods of treatment. U.S. Patent Pending.

Chen WB, Chen JA. (2006). Upright CT Scanner. U.S. Patent 7,003,070.

Chapter 1: Introduction

> The best application of the following beautiful dictum of Geoffroy Saint-Hilaire is in biology: "The infinite is always before us." And the same applies to Carnoy's no less graphic thought: "Science is a perpetual creative process." Not everyone is destined to venture into the forest and by sheer determination carve out a serviceable road. However, even the most humble among us can take advantage of the path opened by genius and by traveling along it extract one or another secret from the unknown.
>
> - Santiago Ramón y Cajal, *Advice for a Young Investigator* – translated by Swanson & Swanson (Ramón y Cajal, 1999)

It is often asked of me, likely out of curiosity more than malicious insinuation, what is the significance of this work? And on the surface one might be hard-pressed to think of it. The quest to explain the genetics of dementia, when we struggle to find a genetic risk factor explaining perhaps 1% of a disease affecting only a handful of people in 100,000, seems perhaps a bit quixotic. Surely a crucial matter for those few people in that minority, but why is this one of the most pressing scientific issues of our time? My reason for believing so is rooted in the uncertainty cloaking the neurodegenerative diseases in general. To work to solve this problem, I keep in mind that scientific discovery, a political event, even a neurodegenerative disease – not just is, but came to be. Consider the brain; consciousness and cognition arise from the chaotic clusters of brainstem nuclei, the hegemony of the higher cortical regions; these in turn are derived from the simple neural tube and its subsequent folding and shaping and thickening throughout development. I learned from Prof. Lazareff that it is through this lens that disease entities like myelomeningoceles and Chiari malformations can be understood (Chen et al., 2012). We have also explored the origin of ideas in about ventriculo-peritoneal (VP) shunting (Cheok et al., 2014) allowing us to evaluate their epistemological rigor. To understand the Chiari malformation we must turn toward the developmental anatomy that set those events into motion. To understand the concept of VP shunting we look to the literature from where the ideas

were developed. But where should we examine for a clearer understanding of the neurodegenerative diseases?

Continuing to trace our capacity to learn and to love, to remember and to desire, through the neural structures, through the migrations of neurons and the flexing and folding of the neural tube, we ultimately arrive at the three billion As, Cs, Gs, and Ts of the human genome. We can now read (mostly) all of these base pairs; but as in the writings of mystics or obscurantists, the task of exegesis is much more daunting. The "genetic code" defines the amino acids encoded by these bases, and we have been able to identify amino acid encoding sequences within "genes" that create proteins, the sequence of which is known and the structures and functions of which are becoming more well-defined. The roughly 20,000 genes themselves, however, clearly fall far short of the lofty functions of the human brain. The remaining 98% of the genome likely accounts for the remainder, but these letters alone are more incomprehensible than the ramblings of Derrida. Metaphorically, but also somewhat literally, we can read the writing of God, but can only understand the nouns – and even among those, we do not know what they mean in most cases. Here, we apply the techniques of modern genetic analysis to the tauopathies – themselves mysterious diseases with intriguing commonalities and differences – ascending Mount Sinai so that we might receive wisdom regarding disease pathogenesis and treatment.

What are tauopathies?

Tau-related pathology can be found across a puzzlingly vast array of neurodegenerative diseases, including some variants of frontotemporal dementia (FTD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), and Alzheimer's disease (AD). Several converging lines of evidence demonstrate that tau plays a causal role in disease pathogenesis: tau pathology often correlates with disease severity and neuronal loss, pathogenic variants in

tau have been found in FTD and PSP, and tau is required for the pathogenesis of many mouse models of Alzheimer's disease.

Abnormal aggregation of tau is present in neurodegenerative diseases with multiple etiologies, including sporadic (PSP, CBD, AD, and argyrophilic grain disease - AGD), genetic (forms of FTD linked to pathogenic tau variants), environmental (chronic traumatic encephalopathy, and perhaps the Parkinson-dementia complex of Guam), and even infectious (the subacute sclerosing pan-encephalitis due to measles). Diseases associated with other pathological protein aggregates, such as AD (with beta amyloid plaques), Parkinson's disease and multiple system atrophy (with predominant Lewy body pathology) and Huntington's disease (with polyglutamine repeats) often have extensive tau pathology (Cairns et al., 1997; Duda et al., 2002; Fernandez-Nogales et al., 2014). Here, we review diseases that have a large body of evidence supporting the central role of tau in their pathogenesis, and in which extensive studies of pathology and genetics have been performed, namely, FTD, PSP, CBD, and AD, focusing particular attention to their pathology at a gross, microscopic, molecular, and ultimately genetic level.

Tauopathies: History and Epidemiology

Frontotemporal Dementia. FTD is a relatively uncommon disease, with incidence estimated at 3-4 per 100,000 person-years in people 45 - 64 years of age, and prevalence at about 15 per 100,000 population (Onyike and Diehl-Schmid, 2013). Because FTD typically presents at an earlier age of onset (mean of about 58 years) in comparison to Alzheimer's disease, it is a common cause of early-onset dementia, with incidence comparable to Alzheimer's disease in the subset of patients under 65 years of age (Johnson et al., 2005; Mercy et al., 2008). In 1892, Pick initially described two cases of patients with fluent aphasia and psychiatric symptoms correlated with focal atrophy of the left temporal lobe, consistent with what is now classified as

FTD (Pick et al., 1994). FTD is a somewhat nebulous concept that includes diseases that may or may not show abnormal tau in neuropathology, and will almost certainly undergo revision as more clearly defined clinico-pathologic correlations are uncovered. Currently, three clinical presentations are considered, each with its characteristic epidemiology, genetic risk factors, and neuropathological features: a behavioral variant, or frontal variant frontotemporal dementia (bvFTD or simply FTD) comprising about 55% of cases, and two language-affecting variants grouped as primary progressive aphasia (PPA), semantic variant PPA (also known as semantic dementia or temporal variant FTD) and nonfluent variant PPA (also known as progressive nonfluent aphasia or agrammatic variant PPA) accounting for 20% and 25% of cases, respectively (Viskontas and Miller, 2007). A logopenic variant PPA is also observed, though it is usually classified as a form of Alzheimer's disease and is associated with $A\beta_{1\text{-}42}$ pathology. Pick's original cases would today be categorized as semantic variant PPA.

FTD presents a relatively loose clinical-neuropathological correlation. In general, tau pathology is characteristic in some cases of bvFTD and in most cases of nonfluent variant PPA (FTD-tau), the latter being closely associated with presentations of CBD and PSP (Josephs et al., 2006). In other cases of bvFTD and in most cases of semantic variant PPA, ubiquitin-positive, tau-negative inclusions that include the protein TDP43 may be observed.

Progressive Supranuclear Palsy. PSP is a relatively rare, rapidly progressing disease, with estimates of prevalence around 5-18 per 100,000 (Coyle-Gilchrist et al., 2016; Nath et al., 2001; Schrag et al., 1999; Takigawa et al., 2016). The disease rarely affects people under age 40, and its incidence ranges from about 1.7 per 100,000 at age 50-59 to 14.7 at age 80-99 years (Bower et al., 1997). PSP was described in nine cases by Clifford Richardson, a neurologist in Toronto, and his colleagues John Steele (then a resident) and Jerzy Olszewski (a neuropathologist) in 1964 (Steele et al., 1964). The group recognized a series of nine cases, with age of onset 50 –

69. These patients exhibited personality changes, vision problems, unsteady gait, mild dementia and behavioral changes, without tremor. Curiously, ophthalmoplegia resulting in vertical gaze palsy was observed; lateral gaze was largely unaffected and full reflexive motion appeared when the head was passively flexed. This unusual sign indicated a supranuclear gaze palsy, in which the neural pathway from the nuclei that control eye movements to the muscles themselves was unaffected, hence leading to the disease's name. Neuropathology demonstrated neurofibrillary tangles, granulovacuolar degeneration, neuronal loss, gliosis, localized to the basal ganglia, brainstem nuclei, and dentate nucleus of the cerebellum.

Even today, those distinctive clinical and pathological signs are still remarkably correlated and homogeneous; the genetics also show remarkable convergence, with almost all cases homozygous for the H1 allele at chromosome 17q21.31. Because the clinical signs are so suggestive of underlying neuropathology, PSP is one of the best "pure tauopathies" (in which tau appears to be the predominant driver of disease) to study the role of tau in neurodegeneration. However, the phenotypic spectrum of PSP is expanding. The classical presentation described by Steele is now termed the "Richardson syndrome"; other presentations include predominant parkinsonism, pure akinesia with gait freezing, corticobasal syndrome, progressive non-fluent aphasia, among others (Respondek and Höglinger, 2016).

Corticobasal Degeneration. The epidemiology of CBD is poorly characterized, though one study has estimated the annual incidence at 0.62-0.92 per 100,000, and prevalence at 4.9-7.3 per 100,000 (Mahapatra et al., 2004). In 1967, Rebeiz and colleagues described three cases with asymmetric "clumsiness", akinesia, ideomotor apraxia, involuntary movements, and upward gaze palsy, which they termed "corticodentatonigral degeneration with neuronal achromasia" (Rebeiz et al., 1968). Cognitive abilities were relatively spared. While the clinical presentation overlapped with PSP, the observations diverged at neuropathological findings of marked

frontoparietal cortical atrophy with swollen and achromatic cortical neurons. This disease entity became known as "corticobasal degeneration" when it was understood that dentate pathology was uncommon. While these pathological findings and the regional distribution of pathology may differ slightly from PSP, the genetics and clinical presentation are highly overlapping and these diseases may be alternative presentations of the same fundamental disease entity; they are often grouped together as so-called "four-repeat tauopathies" (Mailliot et al., 1998).

Alzheimer's disease. AD is by far the most prevalent of the tauopathies, with prevalence in 2017 estimated at 3% in the 65-74 age group, increasing to 16.9% at 75-84 years and 32.1% at 85 years or older, in total affecting 5.3 million people in the United States (Hebert et al., 2013). Incident cases of Alzheimer's disease are expected to reach nearly 500,000 by 2020 (Hebert et al., 2001). In 1906, Alois Alzheimer, a German psychiatrist and neuropathologist, initially described a case of a 51-year old German woman with progressive cognitive impairment with brain atrophy and neurofibrillary tangles; these collective findings were elevated to a distinct disease entity by Kraepelin, which has been revised in the intervening years to what we now refer to as Alzheimer's disease (Berrios, 1990). In 1911, Alzheimer later described a series of cases with plaques and neurofibrillary tangles, further establishing the disease which now bears his name (Alzheimer et al., 1991).

Neurofibrillary tangles were later shown to be composed of the tau protein, and plaques of the beta amyloid protein. However, amyloid pathology appears more central in the disease process, as highly penetrant mutations in amyloid-related proteins (including the amyloid precursor protein, *APP*, and components of amyloid processing complexes, *PSEN1* and *PSEN2*) may cause early onset forms of Alzheimer's disease (Tanzi and Bertram, 2005). Occasionally, tau pathology can appear prior to appreciable amyloid deposition (Crary et al., 2014), although this may simply be a variation of the normal disease process.

The normal biology of tau

Tau was first characterized as a microtubule-binding protein that promoted microtubule polymerization in 1975 (Weingarten et al., 1975); it was not until a decade later that a link with pathological aggregates in Alzheimer's disease was reported (Kosik et al., 1986). Since then, tau and its role in health and disease has been the topic of intense study. The tau protein is widely distributed and specifically expressed throughout the central nervous system. In the adult human brain, tau is expressed in six isoforms resulting from the alternative splicing of exon 10 (one of four tubulin-binding domains) and exons 2 and 3 (the N-terminal inserts). Tau with or without the inclusion of exon 10, termed four-repeat (4R) and three-repeat (3R) tau, is expressed in an approximately 1:1 ratio in healthy adult brain (Hong et al., 1998). The longest of these isoforms (2N 4R) contains 441 amino acids and is used in the nomenclature of tau variants. In the peripheral nervous system, the 2N 4R isoform includes an additional exon, numbered 4a. In fetal brain, the shortest (0N 3R) isoform predominates (Goedert and Jakes, 1990).

Tau knockout mice have also provided insight into the normal function of tau. Surprisingly, initial reports found no overt phenotypes in tau knockout mice; microtubule function appeared slightly affected, and a potentially compensatory upregulation of the microtubule-binding protein MAP1A in early development was observed (Dawson et al., 2001; Fujio et al., 2007; Harada et al., 1994). Mild motor and behavioral deficits have been occasionally observed (Ikegami et al., 2000).

Microtubule binding. Tau was originally described as a microtubule-binding protein, and its effects on microtubule function are well characterized. Exons 9-12 encode four microtubule-binding repeats (18-amino acid sequences) that have high affinity for binding tubulin. Alternative

splicing of exon 10 results in tau isoforms with either three or four of these repeats, with functional differences in microtubule binding affinity. Tau binding promotes microtubule elongation by increasing the rate of polymerization and decreasing transitions to a shrinking state (Drechsel et al., 1992). Binding and unbinding of tau onto tubulin underlies the dynamic instability of microtubules, allowing for remodeling of the cytoskeleton in processes such as axonal growth and neuronal morphogenesis while maintaining microtubule stability.

Neurogenesis. Although neurogenesis in young mice appears preserved, older mice exhibit impaired hippocampal neurogenesis. In one model, new neurons (marked by neuroD and doublecortin) were decreased in tau knockout mice (Hong et al., 2010); in another, the number of neurons were comparable but the localization differed in tau knockout mice (Fuster-Matanzo et al., 2009). Additionally, proliferating neural progenitors in rodent neurogenesis express alternative isoforms of tau (Bullmann et al., 2007). A mouse model of tauopathy expressing human tau also showed deficits in neurogenesis, resulting from reduced proliferation (Komuro et al., 2015). As dynamic microtubule assembly is important for migration and forming projections over long distances (and indeed, key markers of proliferation such as doublecortin are also microtubule-binding proteins) it stands to reason that tau may play a role in neurogenesis, but so far results have been inconsistent and no clear mechanism has emerged.

Axoplasmic transport. Dixit et al. demonstrate that tau binding to microtubules disrupts kinesin- (by detaching at relatively low concentrations) and dynein- (by reversal at higher concentations) mediated transport, and speculate that tau distribution along microtubules may be a mechanism to regulate axoplasmic transport (Dixit et al., 2008). Because of the increased sensitivity to disruption of anterograde transport, accumulation of tau may increase localization of cargoes to the cell body (Stamer et al., 2002). Tau knockout or overexpression mouse strains, however, do not show changes in axoplasmic transport, perhaps due to compensation by other microtubule-

associated proteins (Yuan et al., 2008). Others have proposed alternative mechanisms of tau-mediated disruption of axoplasmic transport (Wang and Mandelkow, 2016).

Neuronal excitability. Tau knockout in mouse and *Drosophila* models of hyperexcitability appears to ameliorate seizure phenotypes (Holth et al., 2013; Roberson et al., 2007), and reduction by antisense oligonucleotides in chemically-induced seizure mouse models (DeVos et al., 2013). In an APP transgenic model of AD, tau knockout resulted in decreased targeting of the Fyn kinase to the dendrite, leading to reduced seizure susceptibility and improved memory and survival (Ittner et al., 2010). The mechanism may be explained by recruitment of Fyn by the projection domain of dendritic tau, resulting in stabilization of the NMDA receptor/PSD-95 interaction, increasing excitability at the synapse. At the synaptic level, these findings may translate into a role of tau in promoting excitotoxicity at high levels.

Protection from DNA damage. Although tau is primarily localized to axons, it is also found in dendrites and nuclei (Loomis et al., 1990), though it function at those locations is poorly characterized. Nuclear tau may have a role in protecting neuronal DNA and RNA integrity (Sultan et al., 2011; Violet et al., 2014). Nuclear tau aggregates have also been reported in disease (Fernandez-Nogales et al., 2014).

Abnormal tau in disease

Since the time of Alzheimer, filamentous aggregates (that were later revealed to be composed of abnormally hyper-phosphorylated tau) have been observed and correlated with neuronal death (Alzheimer et al., 1991). Tau accumulation in the somatodendritic compartment, as opposed to the axon, is also an early feature of tauopathies. Whether these fibrils represent the toxic species of tau or are simply bystanders reflecting underlying disease processes is still debated (Goedert, 2016). Recent thought-provoking work has clarified this issue in some

regard, mapping out the neuron-to-neuron transmission of tau pathology and the kinetics of tau aggregation.

In its normal state, tau is an intrinsically unstructured and highly soluble protein (Mukrasch et al., 2009; Schweers et al., 1994), though negatively charged compounds such as sulfated glucosaminoglycans (Goedert et al., 1996) have been found to seed filament assembly *in vitro*. Foci of local structure exist, notably at the beginning of repeats 2 and 3 (at the sequences ^{275}VQIINK^{280} and ^{306}VQIVYK^{311}); sequences at these regions have a propensity to form β-structure and thereby promote aggregation (von Bergen et al., 2000). The aggregation of tau is thought to proceed by a seeding or nucleation step, followed by elongation. Recent evidence suggests that seeding by tau aggregates can result in spread of tau pathology, reminiscent of the mechanism of prion propagation. Brain extract from mutant P301S tau-expressing mice injected into the hippocampi of wild-type tau expressing mice induced tau pathology that spread to neighboring regions (Clavaguera et al., 2009). Filamentous tau pathology spread into neighboring areas, such as the fimbria, optic tract, and thalamus, and was composed of tau from the wild-type expressing mouse. The seed-competent fraction appears to be insoluble tau, specifically short fibrils, and not tau oligomers (Jackson et al., 2016). Tau-overexpressing mice can also form tau inclusions characteristic of the respective human diseases after inoculation with brain homogenates from patients with FTD, PSP, CBD, and AD (Clavaguera et al., 2013). Despite clear evidence that tau fibrils are the transmitted species, other work has implicated that tau oligomers may be the toxic form (Lasagna-Reeves et al., 2011; Rocher et al., 2010; SantaCruz et al., 2005).

Spread of tau pathology appears to proceed trans-synaptically, because more spatially distant regions may be affected earlier if they are highly connected (Dujardin et al., 2014; Iba et al., 2013; Liu et al., 2012). Tau may be released by exosomes or taken up in new cells by

endocytosis following cell death (Frost et al., 2009; Saman et al., 2012). Although the precise mechanisms remain to be clearly defined, extracellular tau has become an attractive target of therapeutic intervention in recent years.

Pathology

While the tauopathies are defined by pathological tau aggregation, the actual tau aggregation encompasses distinct regional distributions, microscopic appearances, and molecular differences that correlate with clinical features. The remarkable diversity of tau neuropathology findings provides hints at the pathogenesis of this group of diseases, yet their significance is still unclear.

Selective regional vulnerability of brain regions

Each of the tauopathies is classically associated with a particularly affected region. In fact, FTD is named after such a predilection, with the frontal and temporal cortex most affected. Tau pathology is often concentrated in limbic, paralimbic, and frontotemporal cortex (Zhukareva et al., 2002). PSP is known to target the midbrain and dentate nucleus of the cerebellum, and in fact on MRI the atrophic midbrain appears as a "hummingbird sign" (Graber and Staudinger, 2009). Likewise, selectively vulnerable regions in CBD are also hinted by its nomenclature; neurofibrillary tangles and corresponding atrophy is found in the perirolandic cortex, striatum, globus pallidus, and substrantia nigra (Dickson et al., 2002). In AD, the initial site of pathology appears to be the transentorhinal region and subcortical nuclei, and spread into the hippocampus and cortical association areas (Braak and Braak, 1991). The pattern of atrophy in tauopathies (and other neurodegenerative diseases) appears to reflect functional human neural networks, perhaps a consequence of trans-synaptic spread of tau (Seeley et al., 2009).

Microscopic localization and morphology of tau. Tau pathology has typical specific characteristics on a microscopic level, depending on the disease entity. FTD, which has inconsistent correlation between the clinical presentation and neuropathology, manifests in approximately 40% of patients as Pick's disease (PiD). Large round tau inclusions known as "Pick bodies" and achromatic ballooned neurons known as "Pick cells" are hallmark findings (Zhukareva et al., 2002).

PSP and CBD are characterized by filamentous glial and neuronal tau inclusions. The neuropathological presentation can be stereotyped with characteristic findings unique to each disease, though there is a great deal of overlap and intermediate presentations can be observed (Dickson et al., 2002). In oligodendrocytes, tau inclusions appear as "coiled bodies", and are typically more numerous in PSP. In astrocytes, PSP and CBD diverge with tau lesions described as "tufted astrocytes", in which tau reactivity extends to the cell body and "astrocytic plaques", which variably involve the cellular processes but not the cell body, respectively. In neurons, thread-like tau-immunoreactive cell processes in the neuropil are common (Dickson et al., 2007). A characteristic feature of CBD is the presence of achromatic, ballooned neurons, which lack visible Nissl substance and are immunoreactive for neurofilaments, α-B-crystallin, and variably, ubiquitin and tau. In addition, neurons in affected regions may show globose neurofibrillary tangles in PSP and skein-like inclusions and diffuse cytoplasmic tau immunoreactivity ("pre-tangles") in CBD (Dickson et al., 2007).

AD is characterized neuropathologically by neuritic amyloid plaques composed of β-amyloid, neuronal and extracellular neurofibrillary tangles composed of hyperphosphorylated tau (oftentimes described as "flame-shaped"), neuropil threads, and dystrophic neurites, with characteristic spread through the brain in stages (Braak and Braak, 1991). Extracellular "ghost

tangles" composed of remnant neurofibrillary tangles after the death of the containing cell may be present, in contrast to PSP and CBD.

Tau aggregates can also polymerize in variable configurations, and this morphology varies across the spectrum of tauopathies. In disease, tau filaments can wind around each other, crossing over every 80 nm, with a diameter of 10-20 nm (paired helical filaments); remain relatively straight with a diameter of 15 nm (straight filaments); or loosely twist with a periodicity of 100 nm, with a diameter of 15-30 nm (twisted filaments) (Murray et al., 2014). Paired helical filaments predominate in AD, while straight filaments are present in PSP, CBD, and PiD, with occasional twisted filaments. The structural subunits of paired helical filaments and straight filaments appear similar, and the cause of the structural differences remains poorly understood (Crowther, 1991).

Molecular isoforms of tau. Alternative splicing of exon 10, encoding a microtubule-binding domain, can produce two forms of tau: 4R, with four repeated binding sites, or 3R, with three. In healthy brain, these isoforms exist in an approximately 1:1 ratio; however, this ratio may be disrupted in tauopathies, potentially in either direction (Hong et al., 1998). Some forms of FTD show inclusions of predominantly 3R tau (e.g., Pick bodies), whereas aggregates of PSP and CBD are composed of 4R tau. In Alzheimer's disease, a mix of 3R and 4R tau is present. With the addition of an additional tubulin-binding repeat, 4R tau is more effective at stabilizing microtubules and promoting microtubule polymerization but may also be more prone to aggregation (Goode et al., 2000; Panda et al., 2003). Nucleation and extension leading to tau aggregation was promoted by inclusion of exon 10 (Zhong et al., 2012).

Phosphorylation. Aggregates of tau present in the neurofibrillary tangles of tauopathies are abnormally phosphorylated (Grundke-Iqbal et al., 1986). The tau protein has numerous residues

which can be potentially phosphorylated by a wide range of kinases, including glycogen synthase kinase-3 (GSK-3), cyclin-dependent kinase 5 (CDK5), calcium/calmodulin activated protein kinase II (CaMKII), microtubule affinity-regulated kinase 110 (MARK p110) as just a few examples (Dolan and Johnson, 2010; Iqbal et al., 2016). De-phopshorylation of tau is primarily driven by the activity of protein phosphatase 2A (PP2A) (Liu et al., 2005). In general, phosphorylation of tau (both in and out of the tubulin-binding repeats) decreases its affinity to bind to tubulin and reduces its effectiveness to promote polymerization of microtubules (Biernat et al., 1993; Bramblett et al., 1993; Ding et al., 2006). The liberation of tau from microtubules, in addition to direct effects of site-specific phosphorylation, also promotes aggregation. Hyper-phosphorylation of tau may also affect trafficking and turnover (Litersky and Johnson, 1992; Rodríguez-Martín et al., 2013). Depending on the residue, tau phosphorylation may even have a protective function in Alzheimer's disease (Ittner et al., 2016), among other roles.

O-GlcNAcylation. Some authors have reported that O-Linked β-N-acetylglucosamine (O-GlcNAc) is decreased in tau from diseased brains, and nearly absent from insoluble aggregates (Liu et al., 2004). Addition of O-GlcNAc to the hydroxyl groups of serine and threonine competes with (and thus prevents) phosphorylation (Hart and Akimoto, 2009). It may also suppress or enhance transcription and translation, directly affect trafficking, and slow proteolytic turnover. Tau, along with other structural and microtubule-associated proteins, may be modified by O-GlcNAc (Arnold et al., 1996; Liu et al., 2004), though a recent study has raised some skepticism about the true extent of O-GlcNAcylation (Morris et al., 2015). O-GlcNAcylation itself is regulated by the interplay of two proteins, an O-GlcNAc transferase (OGT) and a glycoside hydrolase (O-GlcNAcase, OGA).

Acetylation. Tau appears to be hyper-acetylated at lysine residues of the microtubule-binding region in AD, PSP, and CBD, but not AGD (Grinberg et al., 2013; Irwin et al., 2012). Tau

present in some forms of FTD may have altered acetylation patterns as well (Cohen et al., 2011). Acetylation of tau prevents its ubiquitination and subsequent degradation by the proteasome (Min et al., 2010) and impairs binding to microtubules, promoting aggregation (Cohen et al., 2011; Trzeciakiewicz et al., 2017). Acetylated tau may also be neurotoxic, disrupting neurotransmitter trafficking at the post-synaptic density (Tracy et al., 2016). Other post-translational modifications, such as methylation and ubiquitination, may compete with acetylation at lysine residues and coordinately participate in regulation of tau, though the mechanisms and functions are unclear (Morris et al., 2015).

Genetics

Pathogenic tau variants. Perhaps the strongest evidence of a direct causal role of tau in neurodegeneration is the discovery of pathogenic tau variants in FTD and Parkinsonism linked to Chromosome 17 (FTDP-17) (Hong et al., 1998; Hutton et al., 1998). Many pathogenic tau variants cause autosomal dominant forms of the disease, and have been implicated in phenotypes resembling bvFTD, PSP, CBD, and AD (Goedert and Jakes, 2005). The majority of are clustered in the repeat-binding domains and the splice sites surrounding the alternatively spliced exon 10 (exon 9, intron 9, exon 10, intron 10, exon 11, and exon 12), but have also been reported in exon 1, exon 2, and exon 13 (Spillantini and Goedert, 2013). Duplications of the tau gene have also been found to cause neurodegenerative tauopathies (Hooli et al., 2014; Rovelet-Lecrux et al., 2010).

Numerous downstream mechanisms have been ascribed to these pathogenic tau variants. One of the earliest described consequences was the alternative splicing of exon 10. Intronic variants near the exon 10 5' splice site were among the earliest identified; RNA preparations from the brains of affected patients showed large increases in tau isoforms that included exon 10, and

exon-trapping assays showed that splicing out of exon 10 was reduced in variant-carrying sequences (Hutton et al., 1998). Exonic variants, including N279K, ΔK280, and L284L, also disrupt splicing regulatory elements and affect exon 10 inclusion (D'Souza et al., 1999). Because pathogenic variants have been described that both increase and decrease the inclusion of tau exon 10, the balance between 3R and 4R tau is likely to maintain normal function.

Another category of pathogenic tau variants results in a reduced ability of tau to bind to microtubules and promotes aggregation. Some of the earliest reported pathogenic variants, including ΔK280 and P301L (exon 10) and V337M (exon 12) and R406W (exon 13), fall into this category (Hasegawa et al., 1998; Hong et al., 1998). Many subsequently discovered pathogenic variants were also found to disrupt binding to microtubules, including R5H and R5L (exon 1) (Hayashi et al., 2002; Poorkaj et al., 2002), A152T (exon 7) (Coppola et al., 2012), K257T (exon 9) (Pickering-Brown et al., 2000), and P332S (exon 11) (Deramecourt et al., 2012). While spread across the length of tau, the majority of these appear to cluster near the sequences ^{275}VQIINK^{280} and ^{306}VQIVYK^{311}, which have a propensity for β-structure and also play a crucial role in microtubule binding; indeed, even nearby variants that are not strictly within the sequences (e.g. P301L) may strengthen the tendency for this structure (Mandelkow and Mandelkow, 2012). The coupling of these functions may explain the close relationship between reduced microtubule binding and aggregation in pathogenic tau variants. Of particular interest is the A152T risk-associated variant appeared to decrease the propensity to form fibrillar tau aggregates but increased the formation of tau oligomers (Coppola et al., 2012), suggesting a potential de-coupling between the two. While the aggregates are widely assumed to spread abnormal tau pathology, reports that the oligomers may have greater toxicity suggests that promoting the formation of either species may have toxic downstream effects. The A152T variant is also unique because of its occurrence on exon 7, its occurrence in the healthy

population (suggesting it is a strong risk factor, but not sufficient to cause disease), and association with AD as well as PSP and bvFTD (Lopez et al., 2017).

A number of other mechanisms have been proposed. Pathogenic tau variants may alter its phosphorylation by increasing the efficiency of brain protein kinases, as in the case of R406W (Alonso et al., 2004), or reduce the efficiency of phosphatases such as PP2A (Goedert et al., 2000). Tau turnover may also be affected; for instance, the V337M and R406W tau variants may impair degradation by calpain I (Yen et al., 1999), while the G389R variant appears to have the opposite effect (Pickering-Brown et al., 2000). Recently, Lopez et al. have found that the A152T variant appears to impair proteasome activity, leading to the accumulation of tau and other proteins (Lopez et al., 2017).

Tau haplotype. Conrad and colleagues identified a strong association of allele of a dinucleotide repeat sequence within the *MAPT* gene, which they termed "A0", with PSP (Conrad et al., 1997). While A0/A0 was the major genotype in controls (57.4%), it was highly overrepresented in PSP (95.5%). Later studies reproduced this finding and identified an extended haplotype that explained the association with A0, spanning dozens of genes including *MAPT*. This haplotype resulted from a chromosomal inversion that precluded recombination between the two alleles; H1, tagged by A0, and H2 (Pittman et al., 2004; Steinberg et al., 2012). Aside from PSP, the H1 haplotype is also overrepresented in other diseases, including CBD (Houlden et al., 2001), and curiously, Parkinson's disease (Nalls et al., 2014). While large-scale GWAS in Alzheimer's disease have not identified an association with H1, some large studies have uncovered a potential association with H1 or sub-haplotypes, though the evidence is weak (Allen et al., 2014; Myers et al., 2005; Pastor et al., 2016).

Genes outside of tau – rare variants. Most cases of FTD, PSP, CBD, and AD are sporadic; as previously described, variants in tau itself may occasionally cause tauopathies with overlapping phenotypes. While no gene outside of tau has yet been consistently linked with FTD-tau, PSP, and CBD, several genes have been implicated in familial forms of AD. Notably, *APP*, *PSEN1*, and *PSEN2* are major causes of autosomal dominant, early-onset AD. These genes are each involved in processing β-amyloid, a defining feature of the neuropathology of AD, and may reflect the unique pathogenesis of Alzheimer's disease upstream of tau involvement (Tanzi and Bertram, 2005). The R47H variant of the microglial gene *TREM2* has recently been linked to AD, and may also contribute to some cases of FTD (Guerreiro et al., 2013). Isolated reports of candidate genes abound; for example, pathogenic *PSEN1* variants have been reported in patients with FTD-like tau pathology and without β-amyloid plaques (Dermaut et al., 2004), and an exome array study has highlighted *DYSF* and *PAXIP1* variants in AD (Chen et al., 2015), but more evidence and well-powered sequencing studies are required to forge a definitive link.

Genes outside of tau – common variants from GWAS. As with rare pathogenic variants in familial AD, many polymorphisms have been associated with late-onset AD, because its high prevalence enables the recruitment of large patient cohorts. The ApoE ε4 allele confers an approximately three-fold increase in AD risk for each copy, and is the strongest risk factor identified to date (and is among the strongest common risk factors for any disease) (Corder et al., 1993). Genome-wide association studies have identified variants at a number of loci, including near the genes *CLU*, *CR1*, *MS4A6A*, *ABCA7*, and *BIN1*, among others (Lambert et al., 2013). However, because of primary amyloid pathology, the extent to which these genes participate in tau-related disease processes is difficult to determine.

While the sample sizes are limited in other tauopathies, large-scale genetics studies are beginning to yield new insights. A GWAS of PSP identified robust associations near the genes

STX6, *EIF2AK3*, and *MOBP*, in addition to the H1 haplotype and an independent SNP within the *MAPT* gene, rs242557 (Hoglinger et al., 2011). The strength of these associations, despite the limited sample size, suggests that common variation at a small number of loci plays a relatively large role in the genetic architecture of the disease. The *EIF2AK3* gene encodes PERK, a key component of the unfolded protein response that may play a role in tau metabolism. A GWAS of CBD was underpowered to detect independent associations, but did identify genetic overlap with PSP at the *MOBP* and *MAPT* associations (Kouri et al., 2015). Two GWAS of FTD have been performed, identifying *TMEM106B* (Van Deerlin et al., 2010) and the *HLA* locus (Ferrari et al., 2014), but these studies focused on FTD associated with TDP-43 (not tau) inclusions and clinical FTD, respectively. While interesting footholds have been established in the past decade, the genetics of tauopathies are just beginning to be explored. Understanding the interaction of tau and other genetic risk factors will reveal insights into disease pathways in the future.

Tau-centered therapy in neurodegenerative diseases

Therapeutic strategies to ameliorate tau pathology broadly fall into four categories: decreasing tau levels, altering the posttranslational modifications of tau, blocking the aggregation and spread of tau, or stabilizing microtubules to rescue the normal function of tau. However, despite many attempts at reducing tau pathology, none have so far yielded strong evidence of efficacy. As interest in tau has intensified, more tau-targeted therapeutics are entering the development pipeline, with some early signs of promise. Here, we review these mechanisms and examples of drugs in development falling into each category, with emphasis with the therapeutics that are currently in human trials (Alzforum, 2017).

Decreasing tau levels. Demonstration of transmission of tau between cells in a prion-like mechanism (Clavaguera et al., 2009) suggests that clearance of extracellular or intracellular tau may be a straightforward avenue of treatment for tauopathies. Passive immunotherapy

(Boutajangout et al., 2011) and active vaccination (Asuni et al., 2007) against phosphorylated tau have shown efficacy in proof-of-concept studies of tau P301L mice, resulting in decreased tau pathology and improvements in behavioral tests. A key consideration in the design of immunotherapy against the tau protein is the selective targeting of pathological forms of tau, without mounting an immune response against the normal protein.

Passive immunotherapy is a strategy in which antibodies are administered to directly mount an immune response without actively stimulating the patient's own production of antibodies. BMS-986168 and C2N 8E12 are humanized monoclonal antibodies directed against extracellular forms of tau, which are thought to be involved in the propagation of pathological tau aggregation (Alzforum, 2017; Yanamandra et al., 2013). They are currently being studied in trials for the treatment of AD (ClinicalTrials.gov: NCT02880956) and PSP (ClinicalTrials.gov: NCT02294851, NCT02460094, NCT02494024, NCT02985879). Another example, RG7435 selectively targets tau phosphorylated at S422, which has been identified as a potentially pathogenic species (Laetitia et al., 2012).

Active vaccination, in contrast, stimulates the immune system to mount an active response. For example, ACI-35, is a liposome-based vaccine targeted to tau phosphorylated at S396/S400/S404 (the targets of GSK-3), which showed preferential reduction of abnormally phosphorylated tau in the P301L tau transgenic mouse model (Theunis et al., 2013). Another example, AADvac1 consists of a tau peptide sequence derived from a domain that appears crucial to the oligomerization of tau conjugated to the keyhole limpet haemocyanin protein. Early animal studies showed stark reductions in hyperphosphorylated tau, decreased tau oligomerization and pathology, and improvements in sensorimotor function in transgenic rats expressing truncated human tau (Kontsekova et al., 2014); the vaccine is currently being studied in a Phase II trial in patients with mild AD (ClinicalTrials.gov: NCT02579252).

Tau post-translational modifications. Since phosphorylation and other posttranslational modifications appear to play a central role in the pathogenicity and aggregation of tau, drugs that regulate them may disrupt key disease processes. One of the main kinases involved in tau phosphorylation is GSK-3, and GSK-3 inhibitors have been pursued as a potential treatment for tauopathies. For example, lithium has been shown to inhibit GSK-3 and reduce tau phosphorylation in both *in vitro* and *in vivo* models (Hong et al., 1997; Munoz-Montano et al., 1997). Treatment with lithium for bipolar disease appeared to correlate with decreased risk for Alzheimer's disease in a cross-sectional study (Nunes et al., 2007), and early human trials of lithium showed reductions in CSF phospho-tau and slight reductions in cognitive decline in patients with mild cognitive impairment (Forlenza et al., 2011). However, a subsequent study in PSP and CBD patients showed that lithium was poorly tolerated (Galpern, 2010). Another GSK-3 inhibitor, tideglusib, showed decreased tau phosphorylation, decreased amyloid deposition, and rescue of memory deficits in a mutant APP/tau transgenic mouse model (Serenó et al., 2009). However, trials in AD (Lovestone et al., 2015) and PSP (Tolosa et al., 2014) have missed their primary endpoints.

Promoting a competing protein modification, O-GlcNAc, may also control tau phosphorylation. The O-GlcNAcase inhibitor thiamet-G increased O-GlcNAcylation and decreased phosphorylation of tau in cultured neurons in cultured PC-12 cells and in rat cortex and hippocampus (Yuzwa et al., 2008). Thiamet-G may also inhibit tau aggregation independent of effects on phosphorylation (Yuzwa et al., 2012). A more highly blood-brain barrier penetrant O-GlcNAcase inhibitor, ASN-561, also showed reduction in insoluble tau in the P301S tau transgenic mouse model (Permanne et al., 2015).

Recently, controlling tau acetylation has emerged as a promising avenue of treatment. Administration of salsalate in the PS19 transgenic mouse model decreased tau acetylation at K174 and lowered total tau levels, resulting in decreased hippocampal atrophy and rescued memory deficits (Min et al., 2015). An early-stage trial of salsalate in PSP is currently recruiting patients (ClinicalTrials.gov: NCT02422485).

Tau aggregation inhibitors. Methylene blue was identified as an inhibitor of tau aggregation in a solid-phase tau binding assay (Wischik et al., 1996). It was shown to inhibit tau-tau binding, but not tau-tubulin binding. Possible mechanisms for the anti-aggregation effect of methylene blue and its derivatives include induction of autophagy (Congdon et al., 2012), oxidation of tau cysteine residues resulting in a propensity for a monomeric state (Akoury et al., 2013; Crowe et al., 2013), and deacetylation of lysine 280 (Trzeciakiewicz et al., 2017). The latter two mechanisms are more likely responsible for the initially identified effect, as the assays were independent from the cellular context. Phase 2 trials on a formulation of methylene blue, branded as Rember, demonstrated a mild, but statistically significant effect at the 138 mg/day dose in moderate AD, but not at the 69 mg/day or 228 mg/day doses (Wischik et al., 2015). However, at high doses, potential issues with the formulated release of the drug and its absorption in the digestive system may have interfered with its efficacy (Baddeley et al., 2015). LMTX, a reduced derivative of methylene blue with superior solubility and absorption, was studied in a subsequent Phase 3 trial, which failed to meet its primary endpoints, but follow-up studies continue on these candidates (Gauthier et al., 2016).

Microtubule stabilizers. Physiological tau plays a key role in microtubule stabilization and polymerization, and tau pathology may disrupt this process. Therefore, microtubule-stabilizing agents may ameliorate potential deficits in these pathways in tauopathies. Davunetide is the acetate salt of a short peptide (NAPVSIPQ, also known as NAP) identified as an active fragment

of activity-dependent neuroprotective protein (Bassan et al., 1999). It was initially shown to protect against various neurotoxic insults, such as NMDA toxicity and beta amyloid toxicity. Later work found that NAP stabilized microtubules, promoted microtubule assembly, and reduced tau phosphorylation, reducing tau pathology in rodent tauopathy models (Shiryaev et al., 2009). However, a Phase 2/3 trial of davunetide in PSP did not identify any disease-modifying effect (Boxer et al., 2014). A microtubule stabilizing drug, Epothilone D, was found to be blood-brain barrier penetrant and reduced cognitive deficits in a tauopathy mouse model (Brunden et al., 2010), but development has been discontinued after a Phase I trial (ClinicalTrials.gov: NCT01492374). TPI-287, a blood-brain barrier penetrant taxane that binds tubulin and stabilizes microtubules, has been tested in Phase 1 studies of mild to moderate AD (ClinicalTrials.gov: NCT01966666) and PSP and CBD (ClinicalTrials.gov: NCT02133846).

Rationale

A convergence of evidence places tau as a necessary component of these diseases, including the fact that: pathogenic variants in *MAPT* (the gene encoding tau) have been found in familial forms of FTD linked to chromosome 17; haplotype variants at the 17q21.31 locus encompassing *MAPT* confer risk for PSP, CBD, and Parkinson's Disease; and tau-depleted neurons do not show degeneration after amyloid aggregation. The central role of tau in neurodegeneration has been demonstrated repeatedly, and is gaining even more steam with the widespread failure of alternative targets in Alzheimer's disease. However, the normal biology and its role in disease is still controversial. For example, it is not certain 1) what genetic variants at the 17q21.31 locus increase the risk for neurodegeneration (and how they do so), 2) how genetic variation outside the *MAPT* locus modulates the risk for tauopathy, and ultimately 3) how tau affects disease-related pathways. These gaps in knowledge hinder the development of therapeutic agents that target tau-related pathways, a promising avenue of treatment for neurodegenerative diseases.

To bridge this divide, I investigated the genetic basis of tauopathies using systematic genetic studies and bioinformatics approaches, with the aim of mapping risk genes and understand their downstream consequences. I systematically searched the genome space, including common genetic variation, copy number variants, and rare variants for new disease genes. A genome-wide association study for PSP (Chapter 2) identified novel risk factors and new pathogenic variants, providing a direct causal link to tauopathies. An exome array study for AD, FTD, and PSP (Chapter 3) provided a glimpse into previously uncharted regions of the genome, unearthing yet more candidate disease genes and forging the methods for the application of this new technology. For an ultimate look at (nearly) the entire genome, an ongoing genome sequencing study (Chapter 4) takes a step forward for the discovery of new disease genes. I also examined epigenetic modifications such as DNA methylation (Chapter 5), and described how they relate to genetic risk factors. In Chapter 7, I describe future directions and chart the translation of the work to drug discovery, envisioning how these findings might benefit patients suffering from this devastating group of diseases.

References

Akoury, E., Pickhardt, M., Gajda, M., Biernat, J., Mandelkow, E., and Zweckstetter, M. (2013). Mechanistic Basis of Phenothiazine-Driven Inhibition of Tau Aggregation. Angewandte Chemie International Edition *52*, 3511-3515.

Allen, M., Kachadoorian, M., Quicksall, Z., Zou, F., Chai, H.S., Younkin, C., Crook, J.E., Pankratz, V.S., Carrasquillo, M.M., Krishnan, S.*, et al.* (2014). Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels. Alzheimer's Research and Therapy *6*, 39.

Alonso, A.d.C., Mederlyova, A., Novak, M., Grundke-Iqbal, I., and Iqbal, K. (2004). Promotion of Hyperphosphorylation by Frontotemporal Dementia Tau Mutations. Journal of Biological Chemistry *279*, 34873-34881.

Alzforum (2017). Alzforum Therapeutics Database (Biomedical Research Forum, LLC: accessed March 15, 2017).

Alzheimer, A., Förstl, H., and Levy, R. (1991). On certain peculiar diseases of old age. History of Psychiatry *2*, 71-101.

Arnold, C.S., Johnson, G.V.W., Cole, R.N., Dong, D.L.-Y., Lee, M., and Hart, G.W. (1996). The Microtubule-associated Protein Tau Is Extensively Modified with O-linked N-acetylglucosamine. Journal of Biological Chemistry *271*, 28741-28744.

Asuni, A.A., Boutajangout, A., Quartermain, D., and Sigurdsson, E.M. (2007). Immunotherapy Targeting Pathological Tau Conformers in a Tangle Mouse Model Reduces Brain Pathology with Associated Functional Improvements. The Journal of Neuroscience *27*, 9115-9129.

Baddeley, T.C., McCaffrey, J., M. D. Storey, J., Cheung, J.K.S., Melis, V., Horsley, D., Harrington, C.R., and Wischik, C.M. (2015). Complex Disposition of Methylthioninium Redox Forms Determines Efficacy in Tau Aggregation Inhibitor Therapy for Alzheimer's Disease. Journal of Pharmacology and Experimental Therapeutics *352*, 110-118.

Bassan, M., Zamostiano, R., Davidson, A., Pinhasov, A., Giladi, E., Perl, O., Bassan, H., Blat, C., Gibney, G., Glazner, G.*, et al.* (1999). Complete Sequence of a Novel Protein Containing a Femtomolar-Activity-Dependent Neuroprotective Peptide. Journal of Neurochemistry *72*, 1283-1293.

Berrios, G.E. (1990). Alzheimer's disease: A conceptual history. International Journal of Geriatric Psychiatry *5*, 355-365.

Biernat, J., Gustke, N., Drewes, G., Mandelkow, E., and Mandelkow, E. (1993). Phosphorylation of Ser262 strongly reduces binding of tau to microtubules: Distinction between PHF-like immunoreactivity and microtubule binding. Neuron *11*, 153-163.

Boutajangout, A., Ingadottir, J., Davies, P., and Sigurdsson, E.M. (2011). Passive immunization targeting pathological phospho-tau protein in a mouse model reduces functional decline and clears tau aggregates from the brain. Journal of Neurochemistry *118*, 658-667.

Bower, J.H., Maraganore, D.M., McDonnell, S.K., and Rocca, W.A. (1997). Incidence of progressive supranuclear palsy and multiple system atrophy in Olmsted County, Minnesota, 1976 to 1990. Neurology *49*, 1284-1288.

Boxer, A.L., Lang, A.E., Grossman, M., Knopman, D.S., Miller, B.L., Schneider, L.S., Doody, R.S., Lees, A., Golbe, L.I., Williams, D.R.*, et al.* (2014). Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. The Lancet Neurology *13*, 676-685.

Braak, H., and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. Acta Neuropathologica *82*, 239-259.

Bramblett, G.T., Goedert, M., Jakes, R., Merrick, S.E., Trojanowski, J.Q., and Lee, V.M.Y. (1993). Abnormal tau phosphorylation at Ser396 in alzheimer's disease recapitulates development and contributes to reduced microtubule binding. Neuron *10*, 1089-1099.

Brunden, K.R., Zhang, B., Carroll, J., Yao, Y., Potuzak, J.S., Hogan, A.-M.L., Iba, M., James, M.J., Xie, S.X., Ballatore, C.*, et al.* (2010). Epothilone D Improves Microtubule Density, Axonal Integrity, and Cognition in a Transgenic Mouse Model of Tauopathy. The Journal of Neuroscience *30*, 13861-13866.

Bullmann, T., de Silva, R., Holzer, M., Mori, H., and Arendt, T. (2007). Expression of embryonic tau protein isoforms persist during adult neurogenesis in the hippocampus. Hippocampus *17*, 98-102.

Cairns, N.J., Atkinson, P.F., Hanger, D.P., Anderton, B.H., Daniel, S.E., and L. Lantos, P. (1997). Tau protein in the glial cytoplasmic inclusions of multiple system atrophy can be distinguished from abnormal tau in Alzheimer's disease. Neuroscience Letters *230*, 49-52.

Chen, J., Coutin-Churchman, P., Nuwer, M., and Lazareff, J. (2012). Suboccipital craniotomy for Chiari I results in evoked potential conduction changes. Surgical neurology international *3*, 165.

Chen, J.A., Wang, Q., Davis-Turak, J., and et al. (2015). A multiancestral genome-wide exome array study of alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy. JAMA Neurology *72*, 414-422.

Cheok, S., Chen, J., and Lazareff, J. (2014). The truth and coherence behind the concept of overdrainage of cerebrospinal fluid in hydrocephalic patients. Child's Nervous System *30*, 599-606.

Clavaguera, F., Akatsu, H., Fraser, G., Crowther, R.A., Frank, S., Hench, J., Probst, A., Winkler, D.T., Reichwald, J., Staufenbiel, M.*, et al.* (2013). Brain homogenates from human tauopathies induce tau inclusions in mouse brain. Proceedings of the National Academy of Sciences *110*, 9535-9540.

Clavaguera, F., Bolmont, T., Crowther, R.A., Abramowski, D., Frank, S., Probst, A., Fraser, G., Stalder, A.K., Beibel, M., Staufenbiel, M.*, et al.* (2009). Transmission and spreading of tauopathy in transgenic mouse brain. Nature Cell Biology *11*, 909-913.

Cohen, T.J., Guo, J.L., Hurtado, D.E., Kwong, L.K., Mills, I.P., Trojanowski, J.Q., and Lee, V.M.Y. (2011). The acetylation of tau inhibits its function and promotes pathological tau aggregation. Nature Communications *2*, 252.

Congdon, E.E., Wu, J.W., Myeku, N., Figueroa, Y.H., Herman, M., Marinec, P.S., Gestwicki, J.E., Dickey, C.A., Yu, W.H., and Duff, K.E. (2012). Methylthioninium chloride (methylene blue) induces autophagy and attenuates tauopathy in vitro and in vivo. Autophagy *8*, 609-622.

Conrad, C., Andreadis, A., Trojanowski, J.Q., Dickson, D.W., Kang, D., Chen, X., Wiederholt, W., Hansen, L., Masliah, E., Thal, L.J.*, et al.* (1997). Genetic evidence for the involvement of ? in progressive supranuclear palsy. Annals of Neurology *41*, 277-281.

Coppola, G., Chinnathambi, S., Lee, J.J., Dombroski, B.A., Baker, M.C., Soto-Ortolaza, A.I., Lee, S.E., Klein, E., Huang, A.Y., Sears, R.*, et al.* (2012). Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. Human Molecular Genetics *21*, 3500-3512.

Corder, E., Saunders, A., Strittmatter, W., Schmechel, D., Gaskell, P., Small, G., Roses, A., Haines, J., and Pericak-Vance, M. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science *261*, 921-923.

Coyle-Gilchrist, I.T.S., Dick, K.M., Patterson, K., Vázquez Rodríquez, P., Wehmann, E., Wilcox, A., Lansdall, C.J., Dawson, K.E., Wiggins, J., Mead, S.*, et al.* (2016). Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. Neurology *86*, 1736-1743.

Crary, J.F., Trojanowski, J.Q., Schneider, J.A., Abisambra, J.F., Abner, E.L., Alafuzoff, I., Arnold, S.E., Attems, J., Beach, T.G., Bigio, E.H.*, et al.* (2014). Primary age-related tauopathy (PART): a common pathology associated with human aging. Acta Neuropathologica *128*, 755-766.

Crowe, A., James, M.J., Lee, V.M.-Y., Smith, A.B., Trojanowski, J.Q., Ballatore, C., and Brunden, K.R. (2013). Aminothienopyridazines and Methylene Blue Affect Tau Fibrillization via Cysteine Oxidation. Journal of Biological Chemistry *288*, 11024-11037.

Crowther, R.A. (1991). Straight and paired helical filaments in Alzheimer disease have a common structural unit. Proceedings of the National Academy of Sciences *88*, 2288-2292.

D'Souza, I., Poorkaj, P., Hong, M., Nochlin, D., Lee, V.M.-Y., Bird, T.D., and Schellenberg, G.D. (1999). Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. Proceedings of the National Academy of Sciences *96*, 5598-5603.

Dawson, H.N., Ferreira, A., Eyster, M.V., Ghoshal, N., Binder, L.I., and Vitek, M.P. (2001). Inhibition of neuronal maturation in primary hippocampal neurons from τ deficient mice. Journal of Cell Science *114*, 1179-1187.

Deramecourt, V., Lebert, F., Maurage, C.-A., Fernandez-Gomez, F.-J., Dujardin, S., Colin, M., Sergeant, N., Buée-Scherrer, V., Clot, F., and Ber, I.L. (2012). Clinical, neuropathological, and biochemical characterization of the novel tau mutation P332S. Journal of Alzheimer's Disease *31*, 741-749.

Dermaut, B., Kumar-Singh, S., Engelborghs, S., Theuns, J., Rademakers, R., Saerens, J., Pickut, B.A., Peeters, K., Van Den Broeck, M., Vennekens, K.l*., et al.* (2004). A novel presenilin 1 mutation associated with Pick's disease but not β-amyloid plaques. Annals of Neurology *55*, 617-626.

DeVos, S.L., Goncharoff, D.K., Chen, G., Kebodeaux, C.S., Yamada, K., Stewart, F.R., Schuler, D.R., Maloney, S.E., Wozniak, D.F., Rigo, F*., et al.* (2013). Antisense Reduction of Tau in Adult Mice Protects against Seizures. The Journal of Neuroscience *33*, 12887-12897.

Dickson, D.W., Bergeron, C., Chin, S.S., Duyckaerts, C., Horoupian, D., Ikeda, K., Jellinger, K., Lantos, P.L., Lippa, C.F., Mirra, S.S*., et al.* (2002). Office of Rare Diseases Neuropathologic Criteria for Corticobasal Degeneration. Journal of Neuropathology and Experimental Neurology *61*, 935-946.

Dickson, D.W., Rademakers, R., and Hutton, M.L. (2007). Progressive Supranuclear Palsy: Pathology and Genetics. Brain Pathology *17*, 74-82.

Ding, H., Matthews, T.A., and Johnson, G.V.W. (2006). Site-specific Phosphorylation and Caspase Cleavage Differentially Impact Tau-Microtubule Interactions and Tau Aggregation. Journal of Biological Chemistry *281*, 19107-19114.

Dixit, R., Ross, J.L., Goldman, Y.E., and Holzbaur, E.L.F. (2008). Differential Regulation of Dynein and Kinesin Motor Proteins by Tau. Science *319*, 1086-1089.

Dolan, P.J., and Johnson, G.V. (2010). The role of tau kinases in Alzheimer's disease. Current Opinion in Drug Discovery and Development *13*, 595.

Drechsel, D.N., Hyman, A.A., Cobb, M.H., and Kirschner, M.W. (1992). Modulation of the dynamic instability of tubulin assembly by the microtubule-associated protein tau. Molecular Biology of the Cell *3*, 1141-1154.

Duda, J.E., Giasson, B.I., Mabon, M.E., Miller, D.C., Golbe, L.I., Lee, V.M.-Y., and Trojanowski, J.Q. (2002). Concurrence of α-synuclein and tau brain pathology in the Contursi kindred. Acta Neuropathologica *104*, 7-11.

Dujardin, S., Lécolle, K., Caillierez, R., Bégard, S., Zommer, N., Lachaud, C., Carrier, S., Dufour, N., Aurégan, G., Winderickx, J.*, et al.* (2014). Neuron-to-neuron wild-type Tau protein transfer through a trans-synaptic mechanism: relevance to sporadic tauopathies. Acta Neuropathologica Communications *2*, 14.

Fernandez-Nogales, M., Cabrera, J.R., Santos-Galindo, M., Hoozemans, J.J.M., Ferrer, I., Rozemuller, A.J.M., Hernandez, F., Avila, J., and Lucas, J.J. (2014). Huntington's disease is a four-repeat tauopathy with tau nuclear rods. Nature Medicine *20*, 881-885.

Ferrari, R., Hernandez, D.G., Nalls, M.A., Rohrer, J.D., Ramasamy, A., Kwok, J.B.J., Dobson-Stone, C., Brooks, W.S., Schofield, P.R., Halliday, G.M.*, et al.* (2014). Frontotemporal dementia and its subtypes: a genome-wide association study. The Lancet Neurology *13*, 686-699.

Forlenza, O.V., Diniz, B.S., Radanovic, M., Santos, F.S., Talib, L.L., and Gattaz, W.F. (2011). Disease-modifying properties of long-term lithium treatment for amnestic mild cognitive impairment: randomised controlled trial. The British Journal of Psychiatry *198*, 351-356.

Frost, B., Jacks, R.L., and Diamond, M.I. (2009). Propagation of Tau Misfolding from the Outside to the Inside of a Cell. Journal of Biological Chemistry *284*, 12845-12852.

Fujio, K., Sato, M., Uemura, T., Sato, T., Sato-Harada, R., and Harada, A. (2007). 14-3-3 proteins and protein phosphatases are not reduced in tau-deficient mice. Neuroreport *18*, 1049-1052.

Fuster-Matanzo, A., de Barreda, E.G., Dawson, H.N., Vitek, M.P., Avila, J., and Hernández, F. (2009). Function of tau protein in adult newborn neurons. FEBS Letters *583*, 3063-3068.

Galpern, W. (2010). Lithium in progressive supranuclear palsy and corticobasal degeneration (Abstracts of the Fourteenth International Congress of Parkinson's Disease and Movement Disorders). Movement Disorders *25*, S498.

Gauthier, S., Feldman, H.H., Schneider, L.S., Wilcock, G.K., Frisoni, G.B., Hardlund, J.H., Moebius, H.J., Bentham, P., Kook, K.A., Wischik, D.J.*, et al.* (2016). Efficacy and safety of tau-aggregation inhibitor therapy in patients with mild or moderate Alzheimer's disease: a randomised, controlled, double-blind, parallel-arm, phase 3 trial. The Lancet *388*, 2873-2884.

Goedert, M. (2016). The ordered assembly of tau is the gain-of-toxic function that causes human tauopathies. Alzheimer's & Dementia *12*, 1040-1050.

Goedert, M., and Jakes, R. (1990). Expression of separate isoforms of human tau protein: correlation with the tau pattern in brain and effects on tubulin polymerization. The EMBO journal *9*, 4225-4230.

Goedert, M., and Jakes, R. (2005). Mutations causing neurodegenerative tauopathies. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease *1739*, 240-250.

Goedert, M., Jakes, R., Spillantini, M.G., Hasegawa, M., Smith, M.J., and Crowther, R.A. (1996). Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans. Nature *383*, 550-553.

Goedert, M., Satumtira, S., Jakes, R., Smith, M.J., Kamibayashi, C., White, C.L., and Sontag, E. (2000). Reduced Binding of Protein Phosphatase 2A to Tau Protein with Frontotemporal Dementia and Parkinsonism Linked to Chromosome 17 Mutations. Journal of Neurochemistry *75*, 2155-2162.

Goode, B.L., Chau, M., Denis, P.E., and Feinstein, S.C. (2000). Structural and Functional Differences between 3-Repeat and 4-Repeat Tau Isoforms: Implications for normal tau function and the onset of neurodegenerative disease. Journal of Biological Chemistry *275*, 38182-38189.

Graber, J.J., and Staudinger, R. (2009). Teaching NeuroImages: “Penguin” or “hummingbird” sign and midbrain atrophy in progressive supranuclear palsy. Neurology *72*, e81.

Grinberg, L.T., Wang, X., Wang, C., Sohn, P.D., Theofilas, P., Sidhu, M., Arevalo, J.B., Heinsen, H., Huang, E.J., Rosen, H.*, et al.* (2013). Argyrophilic grain disease differs from other tauopathies by lacking tau acetylation. Acta Neuropathologica *125*, 581-593.

Grundke-Iqbal, I., Iqbal, K., Tung, Y.C., Quinlan, M., Wisniewski, H.M., and Binder, L.I. (1986). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. Proceedings of the National Academy of Sciences *83*, 4913-4917.

Guerreiro, R., Lohmann, E., Brás, J., and et al. (2013). Using exome sequencing to reveal mutations in trem2 presenting as a frontotemporal dementia–like syndrome without bone involvement. JAMA Neurology *70*, 78-84.

Harada, A., Oguchi, K., Okabe, S., Kuno, J., Terada, S., Ohshima, T., Sato-Yoshitake, R., Takei, Y., Noda, T., and Hirokawa, N. (1994). Altered microtubule organization in small-calibre axons of mice lacking tau protein. Nature *369*, 488-491.

Hart, G.W., and Akimoto, Y. (2009). The O-GlcNAc Modification. In Essentials of Glycobiology, 2nd edition, A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.R. Hart, and M.E. Etzler, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Hasegawa, M., Smith, M.J., and Goedert, M. (1998). Tau proteins with FTDP‐17 mutations have a reduced ability to promote microtubule assembly. FEBS Letters *437*, 207-210.

Hayashi, S., Toyoshima, Y., Hasegawa, M., Umeda, Y., Wakabayashi, K., Tokiguchi, S., Iwatsubo, T., and Takahashi, H. (2002). Late-onset frontotemporal dementia with a novel exon 1 (Arg5His) tau gene mutation. Annals of Neurology *51*, 525-530.

Hebert, L.E., Beckett, L.A., Scherr, P.A., and Evans, D.A. (2001). Annual Incidence of Alzheimer Disease in the United States Projected to the Years 2000 Through 2050. Alzheimer Disease and Associated Disorders *15*, 169-173.

Hebert, L.E., Weuve, J., Scherr, P.A., and Evans, D.A. (2013). Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology *80*, 1778-1783.

Hoglinger, G.U., Melhem, N.M., Dickson, D.W., Sleiman, P.M.A., Wang, L.-S., Klei, L., Rademakers, R., de Silva, R., Litvan, I., Riley, D.E.*, et al.* (2011). Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nature Genetics *43*, 699-705.

Holth, J.K., Bomben, V.C., Reed, J.G., Inoue, T., Younkin, L., Younkin, S.G., Pautler, R.G., Botas, J., and Noebels, J.L. (2013). Tau Loss Attenuates Neuronal Network Hyperexcitability in Mouse and Drosophila Genetic Models of Epilepsy. The Journal of Neuroscience *33*, 1651-1659.

Hong, M., Chen, D.C.R., Klein, P.S., and Lee, V.M.-Y. (1997). Lithium Reduces Tau Phosphorylation by Inhibition of Glycogen Synthase Kinase-3. Journal of Biological Chemistry *272*, 25326-25332.

Hong, M., Zhukareva, V., Vogelsberg-Ragaglia, V., Wszolek, Z., Reed, L., Miller, B.I., Geschwind, D.H., Bird, T.D., McKeel, D., Goate, A.*, et al.* (1998). Mutation-Specific Functional Impairments in Distinct Tau Isoforms of Hereditary FTDP-17. Science *282*, 1914-1917.

Hong, X.-P., Peng, C.-X., Wei, W., Tian, Q., Liu, Y.-H., Yao, X.-Q., Zhang, Y., Cao, F.-Y., Wang, Q., and Wang, J.-Z. (2010). Essential role of tau phosphorylation in adult hippocampal neurogenesis. Hippocampus *20*, 1339-1349.

Hooli, B.V., Kovacs-Vajna, Z.M., Mullin, K., Blumenthal, M.A., Mattheisen, M., Zhang, C., Lange, C., Mohapatra, G., Bertram, L., and Tanzi, R.E. (2014). Rare autosomal copy number variations in early-onset familial Alzheimer/'s disease. Molecular Psychiatry *19*, 676-681.

Houlden, H., Baker, M., Morris, H.R., MacDonald, N., Pickering–Brown, S., Adamson, J., Lees, A.J., Rossor, M.N., Quinn, N.P., Kertesz, A.*, et al.* (2001). Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype. Neurology *56*, 1702-1706.

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A.*, et al.* (1998). Association of missense and 5[prime]-splice-site mutations in tau with the inherited dementia FTDP-17. Nature *393*, 702-705.

Iba, M., Guo, J.L., McBride, J.D., Zhang, B., Trojanowski, J.Q., and Lee, V.M.-Y. (2013). Synthetic Tau Fibrils Mediate Transmission of Neurofibrillary Tangles in a Transgenic Mouse Model of Alzheimer's-Like Tauopathy. The Journal of Neuroscience *33*, 1024-1037.

Ikegami, S., Harada, A., and Hirokawa, N. (2000). Muscle weakness, hyperactivity, and impairment in fear conditioning in tau-deficient mice. Neuroscience Letters *279*, 129-132.

Iqbal, K., Liu, F., and Gong, C.-X. (2016). Tau and neurodegenerative disease: the story so far. Nat Rev Neurol *12*, 15-27.

Irwin, D.J., Cohen, T.J., Grossman, M., Arnold, S.E., Xie, S.X., Lee, V.M.Y., and Trojanowski, J.Q. (2012). Acetylated tau, a novel pathological signature in Alzheimer's disease and other tauopathies. Brain *135*, 807-818.

Ittner, A., Chua, S.W., Bertz, J., Volkerling, A., van der Hoven, J., Gladbach, A., Przybyla, M., Bi, M., van Hummel, A., Stevens, C.H.*, et al.* (2016). Site-specific phosphorylation of tau inhibits amyloid-β toxicity in Alzheimer's mice. Science *354*, 904-908.

Ittner, L.M., Ke, Y.D., Delerue, F., Bi, M., Gladbach, A., van Eersel, J., Wölfing, H., Chieng, B.C., Christie, M.J., Napier, I.A.*, et al.* (2010). Dendritic Function of Tau Mediates Amyloid-β Toxicity in Alzheimer's Disease Mouse Models. Cell *142*, 387-397.

Jackson, S.J., Kerridge, C., Cooper, J., Cavallini, A., Falcon, B., Cella, C.V., Landi, A., Szekeres, P.G., Murray, T.K., Ahmed, Z.*, et al.* (2016). Short Fibrils Constitute the Major Species of Seed-Competent Tau in the Brains of Mice Transgenic for Human P301S Tau. The Journal of Neuroscience *36*, 762-772.

Johnson, J.K., Diehl, J., Mendez, M.F., and et al. (2005). Frontotemporal lobar degeneration: Demographic characteristics of 353 patients. Archives of Neurology *62*, 925-930.

Josephs, K.A., Petersen, R.C., Knopman, D.S., Boeve, B.F., Whitwell, J.L., Duffy, J.R., Parisi, J.E., and Dickson, D.W. (2006). Clinicopathologic analysis of frontotemporal and corticobasal degenerations and PSP. Neurology *66*, 41-48.

Komuro, Y., Xu, G., Bhaskar, K., and Lamb, B.T. (2015). Human tau expression reduces adult neurogenesis in a mouse model of tauopathy. Neurobiology of Aging *36*, 2034-2042.

Kontsekova, E., Zilka, N., Kovacech, B., Novak, P., and Novak, M. (2014). First-in-man tau vaccine targeting structural determinants essential for pathological tau–tau interaction reduces tau oligomerisation and neurofibrillary degeneration in an Alzheimer's disease model. Alzheimer's Research and Therapy *6*, 44.

Kosik, K.S., Joachim, C.L., and Selkoe, D.J. (1986). Microtubule-associated protein tau (tau) is a major antigenic component of paired helical filaments in Alzheimer disease. Proceedings of the National Academy of Sciences *83*, 4044-4048.

Kouri, N., Ross, O.A., Dombroski, B., Younkin, C.S., Serie, D.J., Soto-Ortolaza, A., Baker, M., Finch, N.C.A., Yoon, H., Kim, J.*, et al.* (2015). Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. Nature Communications *6*, 7247.

Laetitia, T., Raphaelle, C., Sylvie, B., Francisco, J.F.-G., Marie-Eve, G., Nadege, Z., Nicolas, S., Susanna, S.-M., David, B., and Luc, B. (2012). Targeting Phospho-Ser422 by Active Tau Immunotherapy in the THYTau22 Mouse Model: A Suitable Therapeutic Approach. Current Alzheimer Research *9*, 397-405.

Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W.*, et al.* (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature Genetics *45*, 1452-1458.

Lasagna-Reeves, C.A., Castillo-Carranza, D.L., Sengupta, U., Clos, A.L., Jackson, G.R., and Kayed, R. (2011). Tau oligomers impair memory and induce synaptic and mitochondrial dysfunction in wild-type mice. Molecular Neurodegeneration *6*, 39.

Litersky, J.M., and Johnson, G.V. (1992). Phosphorylation by cAMP-dependent protein kinase inhibits the degradation of tau by calpain. Journal of Biological Chemistry *267*, 1563-1568.

Liu, F., Grundke-Iqbal, I., Iqbal, K., and Gong, C.-X. (2005). Contributions of protein phosphatases PP1, PP2A, PP2B and PP5 to the regulation of tau phosphorylation. European Journal of Neuroscience *22*, 1942-1950.

Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G.W., and Gong, C.-X. (2004). O-GlcNAcylation regulates phosphorylation of tau: A mechanism involved in Alzheimer's disease. Proceedings of the National Academy of Sciences of the United States of America *101*, 10804-10809.

Liu, L., Drouet, V., Wu, J.W., Witter, M.P., Small, S.A., Clelland, C., and Duff, K. (2012). Trans-Synaptic Spread of Tau Pathology In Vivo. PLoS ONE *7*, e31302.

Loomis, P.A., Howard, T.H., Castleberry, R.P., and Binder, L.I. (1990). Identification of nuclear tau isoforms in human neuroblastoma cells. Proceedings of the National Academy of Sciences *87*, 8422-8426.

Lopez, A., Lee, S.E., Wojta, K., Ramos, E.M., Klein, E., Chen, J., Boxer, A.L., Gorno-Tempini, M.L., Geschwind, D.H., Schlotawa, L.*, et al.* (2017). A152T tau allele causes neurodegeneration that can be ameliorated in a zebrafish model by autophagy induction. Brain *140*, 1128-1146.

Lovestone, S., Boada, M., Dubois, B., Hüll, M., Rinne, J.O., Huppertz, H.-J., Calero, M., Andrés, M.V., Gómez-Carrillo, B., and León, T. (2015). A phase II trial of tideglusib in Alzheimer's disease. Journal of Alzheimer's Disease *45*, 75-88.

Mahapatra, R.K., Edwards, M.J., Schott, J.M., and Bhatia, K.P. (2004). Corticobasal degeneration. The Lancet Neurology *3*, 736-743.

Mailliot, C., Sergeant, N., Bussière, T., Caillet-Boudin, M.-L., Delacourte, A., and Buée, L. (1998). Phosphorylation of specific sets of tau isoforms reflects different neurofibrillary degeneration processes. FEBS Letters *433*, 201-204.

Mandelkow, E.-M., and Mandelkow, E. (2012). Biochemistry and Cell Biology of Tau Protein in Neurofibrillary Degeneration. Cold Spring Harbor Perspectives in Medicine *2*.

Mercy, L., Hodges, J.R., Dawson, K., Barker, R.A., and Brayne, C. (2008). Incidence of early-onset dementias in Cambridgeshire, United Kingdom. Neurology *71*, 1496-1499.

Min, S.-W., Chen, X., Tracy, T.E., Li, Y., Zhou, Y., Wang, C., Shirakawa, K., Minami, S.S., Defensor, E., Mok, S.A.*, et al.* (2015). Critical role of acetylation in tau-mediated neurodegeneration and cognitive deficits. Nature Medicine *21*, 1154-1162.

Min, S.-W., Cho, S.-H., Zhou, Y., Schroeder, S., Haroutunian, V., Seeley, W.W., Huang, E.J., Shen, Y., Masliah, E., Mukherjee, C.*, et al.* (2010). Acetylation of Tau Inhibits Its Degradation and Contributes to Tauopathy. Neuron *67*, 953-966.

Morris, M., Knudsen, G.M., Maeda, S., Trinidad, J.C., Ioanoviciu, A., Burlingame, A.L., and Mucke, L. (2015). Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. Nature Neuroscience *18*, 1183-1189.

Mukrasch, M.D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2009). Structural Polymorphism of 441-Residue Tau at Single Residue Resolution. PLoS Biology *7*, e1000034.

Munoz-Montano, J.R., Moreno, F.J., Avila, J., and Diaz-Nido, J. (1997). Lithium inhibits Alzheimer's disease-like tau protein phosphorylation in neurons. FEBS Letters *411*, 183-188.

Murray, M.E., Kouri, N., Lin, W.-L., Jack, C.R., Dickson, D.W., and Vemuri, P. (2014). Clinicopathologic assessment and imaging of tauopathies in neurodegenerative dementias. Alzheimer's Research and Therapy *6*, 1.

Myers, A.J., Kaleem, M., Marlowe, L., Pittman, A.M., Lees, A.J., Fung, H.C., Duckworth, J., Leung, D., Gibson, A., Morris, C.M.*, et al.* (2005). The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. Human Molecular Genetics *14*, 2399-2404.

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M.*, et al.* (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nature Genetics *46*, 989-993.

Nath, U., Ben-Shlomo, Y., Thomson, R.G., Morris, H.R., Wood, N.W., Lees, A.J., and Burn, D.J. (2001). The prevalence of progressive supranuclear palsy (Steele–Richardson–Olszewski syndrome) in the UK. Brain *124*, 1438-1449.

Nunes, P.V., Forlenza, O.V., and Gattaz, W.F. (2007). Lithium and risk for Alzheimer's disease in elderly patients with bipolar disorder. The British Journal of Psychiatry *190*, 359-360.

Onyike, C.U., and Diehl-Schmid, J. (2013). The epidemiology of frontotemporal dementia. International Review of Psychiatry *25*, 130-137.

Panda, D., Samuel, J.C., Massie, M., Feinstein, S.C., and Wilson, L. (2003). Differential regulation of microtubule dynamics by three- and four-repeat tau: Implications for the onset of neurodegenerative disease. Proceedings of the National Academy of Sciences *100*, 9548-9553.

Pastor, P., Moreno, F., Clarimón, J., Ruiz, A., Combarros, O., Calero, M., de Munain, A.L., Bullido, M.J., de Pancorbo, M.M., and Carro, E. (2016). MAPT H1 haplotype is associated with late-onset alzheimer's disease risk in APOEε4 noncarriers: results from the Dementia Genetics Spanish Consortium. Journal of Alzheimer's Disease *49*, 343-352.

Permanne, B., Quattropani, A., Hantson, J., Neny, M., Ousson, S., Sand, A., Wiessner, C., and Beher, D. (2015). Pharmacological intervention with the novel o-glcnacase inhibitor ASN-561 reduces pathological tau in transgenic mice. Alzheimer's & Dementia: The Journal of the Alzheimer's Association *11*, P227.

Pick, A., Girling, D.M., and Berrios, G.E. (1994). On the relationship between senile cerebral atrophy and aphasia. History of Psychiatry *5*, 542-547.

Pickering-Brown, S., Baker, M., Yen, S.-H., Liu, W.-K., Hasegawa, M., Cairns, N., Lantos, P.L., Rossor, M., Iwatsubo, T., and Davies, Y. (2000). Pick's disease is associated with mutations in the tau gene. Annals of Neurology *48*, 859-867.

Pittman, A.M., Myers, A.J., Duckworth, J., Bryden, L., Hanson, M., Abou-Sleiman, P., Wood, N.W., Hardy, J., Lees, A., and de Silva, R. (2004). The structure of the tau haplotype in controls and in progressive supranuclear palsy. Human Molecular Genetics *13*, 1267-1274.

Poorkaj, P., Muma, N.A., Zhukareva, V., Cochran, E.J., Shannon, K.M., Hurtig, H., Koller, W.C., Bird, T.D., Trojanowski, J.Q., Lee, V.M.Y.*, et al.* (2002). An R5L τ mutation in a subject with a progressive supranuclear palsy phenotype. Annals of Neurology *52*, 511-516.

Ramón y Cajal, S. (1999). Advice for a Young Investigator (Cambridge, MA: The MIT Press).

Rebeiz, J.J., Kolodny, E.H., Richardson, E.P., and Jr (1968). Corticodentatonigral degeneration with neuronal achromasia. Archives of Neurology *18*, 20-33.

Respondek, G., and Höglinger, G.U. (2016). The phenotypic spectrum of progressive supranuclear palsy. Parkinsonism & Related Disorders *22, Supplement 1*, S34-S36.

Roberson, E.D., Scearce-Levie, K., Palop, J.J., Yan, F., Cheng, I.H., Wu, T., Gerstein, H., Yu, G.-Q., and Mucke, L. (2007). Reducing Endogenous Tau Ameliorates Amyloid ß-Induced Deficits in an Alzheimer's Disease Mouse Model. Science *316*, 750-754.

Rocher, A.B., Crimins, J.L., Amatrudo, J.M., Kinson, M.S., Todd-Brown, M.A., Lewis, J., and Luebke, J.I. (2010). Structural and functional changes in tau mutant mice neurons are not linked to the presence of NFTs. Experimental Neurology *223*, 385-393.

Rodríguez-Martín, T., Cuchillo-Ibáñez, I., Noble, W., Nyenya, F., Anderton, B.H., and Hanger, D.P. (2013). Tau phosphorylation affects its axonal transport and degradation. Neurobiology of Aging *34*, 2146-2157.

Rovelet-Lecrux, A., Hannequin, D., Guillin, O., Legallic, S., Jurici, S., Wallon, D., Frebourg, T., and Campion, D. (2010). Frontotemporal dementia phenotype associated with MAPT gene duplication. Journal of Alzheimer's Disease *21*, 897-902.

Saman, S., Kim, W., Raya, M., Visnick, Y., Miro, S., Saman, S., Jackson, B., McKee, A.C., Alvarez, V.E., Lee, N.C.Y.*, et al.* (2012). Exosome-associated Tau Is Secreted in Tauopathy Models and Is Selectively Phosphorylated in Cerebrospinal Fluid in Early Alzheimer Disease. Journal of Biological Chemistry *287*, 3842-3849.

SantaCruz, K., Lewis, J., Spires, T., Paulson, J., Kotilinek, L., Ingelsson, M., Guimaraes, A., DeTure, M., Ramsden, M., McGowan, E.*, et al.* (2005). Tau Suppression in a Neurodegenerative Mouse Model Improves Memory Function. Science *309*, 476-481.

Schrag, A., Ben-Shlomo, Y., and Quinn, N.P. (1999). Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. The Lancet *354*, 1771-1775.

Schweers, O., Schönbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994). Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. Journal of Biological Chemistry *269*, 24290-24297.

Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L., and Greicius, M.D. (2009). Neurodegenerative Diseases Target Large-Scale Human Brain Networks. Neuron *62*, 42-52.

Serenó, L., Coma, M., Rodríguez, M., Sánchez-Ferrer, P., Sánchez, M.B., Gich, I., Agulló, J.M., Pérez, M., Avila, J., Guardia-Laguarta, C.*, et al.* (2009). A novel GSK-3β inhibitor reduces Alzheimer's pathology and rescues neuronal loss in vivo. Neurobiology of Disease *35*, 359-367.

Shiryaev, N., Jouroukhin, Y., Giladi, E., Polyzoidou, E., Grigoriadis, N.C., Rosenmann, H., and Gozes, I. (2009). NAP protects memory, increases soluble tau and reduces tau hyperphosphorylation in a tauopathy model. Neurobiology of Disease *34*, 381-388.

Spillantini, M.G., and Goedert, M. (2013). Tau pathology and neurodegeneration. The Lancet Neurology *12*, 609-622.

Stamer, K., Vogel, R., Thies, E., Mandelkow, E., and Mandelkow, E.-M. (2002). Tau blocks traffic of organelles, neurofilaments, and APP vesicles in neurons and enhances oxidative stress. The Journal of Cell Biology *156*, 1051-1063.

Steele, J.C., Richardson, J., and Olszewski, J. (1964). Progressive supranuclear palsy: A heterogeneous degeneration involving the brain stem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. Archives of Neurology *10*, 333-359.

Steinberg, K.M., Antonacci, F., Sudmant, P.H., Kidd, J.M., Campbell, C.D., Vives, L., Malig, M., Scheinfeldt, L., Beggs, W., Ibrahim, M.*, et al.* (2012). Structural diversity and African origin of the 17q21.31 inversion polymorphism. Nature Genetics *44*, 872-880.

Sultan, A., Nesslany, F., Violet, M., Bégard, S., Loyens, A., Talahari, S., Mansuroglu, Z., Marzin, D., Sergeant, N., Humez, S.*, et al.* (2011). Nuclear Tau, a Key Player in Neuronal DNA Protection. Journal of Biological Chemistry *286*, 4566-4575.

Takigawa, H., Kitayama, M., Wada-Isoe, K., Kowa, H., and Nakashima, K. (2016). Prevalence of progressive supranuclear palsy in Yonago: change throughout a decade. Brain and Behavior *6*, e00557-n/a.

Tanzi, R.E., and Bertram, L. (2005). Twenty Years of the Alzheimer's Disease Amyloid Hypothesis: A Genetic Perspective. Cell *120*, 545-555.

Theunis, C., Crespo-Biel, N., Gafner, V., Pihlgren, M., López-Deber, M.P., Reis, P., Hickman, D.T., Adolfsson, O., Chuard, N., Ndao, D.M.*, et al.* (2013). Efficacy and Safety of A Liposome-Based Vaccine against Protein Tau, Assessed in Tau.P301L Mice That Model Tauopathy. PLoS ONE *8*, e72301.

Tolosa, E., Litvan, I., Höglinger, G.U., Burn, D., Lees, A., Andrés, M.V., Gómez-Carrillo, B., León, T., del Ser, T., and for the, T.I. (2014). A phase 2 trial of the GSK-3 inhibitor tideglusib in progressive supranuclear palsy. Movement Disorders *29*, 470-478.

Tracy, Tara E., Sohn, Peter D., Minami, S.S., Wang, C., Min, S.-W., Li, Y., Zhou, Y., Le, D., Lo, I., Ponnusamy, R.*, et al.* (2016). Acetylated Tau Obstructs KIBRA-Mediated Signaling in Synaptic Plasticity and Promotes Tauopathy-Related Memory Loss. Neuron *90*, 245-260.

Trzeciakiewicz, H., Tseng, J.-H., Wander, C.M., Madden, V., Tripathy, A., Yuan, C.-X., and Cohen, T.J. (2017). A Dual Pathogenic Mechanism Links Tau Acetylation to Sporadic Tauopathy. Scientific Reports *7*, 44102.

Van Deerlin, V.M., Sleiman, P.M.A., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.-S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M.*, et al.* (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nature Genetics *42*, 234-239.

Violet, M., Delattre, L., Tardivel, M., Sultan, A., Chauderlier, A., Caillierez, R., Talahari, S., Nesslany, F., Lefebvre, B., Bonnefoy, E.*, et al.* (2014). A major role for Tau in neuronal DNA and RNA protection in vivo under physiological and hyperthermic conditions. Frontiers in Cellular Neuroscience *8*.

Viskontas, I., and Miller, B. (2007). FRONTOTEMPORAL DEMENTIA. CONTINUUM: Lifelong Learning in Neurology *13*, 87-108.

von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E.-M., and Mandelkow, E. (2000). Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif (306VQIVYK311) forming β structure. Proceedings of the National Academy of Sciences *97*, 5129-5134.

Wang, Y., and Mandelkow, E. (2016). Tau in physiology and pathology. Nature Reviews Neuroscience *17*, 22-35.

Weingarten, M.D., Lockwood, A.H., Hwo, S.Y., and Kirschner, M.W. (1975). A protein factor essential for microtubule assembly. Proceedings of the National Academy of Sciences *72*, 1858-1862.

Wischik, C.M., Edwards, P.C., Lai, R.Y., Roth, M., and Harrington, C.R. (1996). Selective inhibition of Alzheimer disease-like tau aggregation by phenothiazines. Proceedings of the National Academy of Sciences *93*, 11213-11218.

Wischik, C.M., Staff, R.T., Wischik, D.J., Bentham, P., Murray, A.D., Storey, J., Kook, K.A., and Harrington, C.R. (2015). Tau aggregation inhibitor therapy: an exploratory phase 2 study in mild or moderate Alzheimer's disease. Journal of Alzheimer's Disease *44*, 705-720.

Yanamandra, K., Kfoury, N., Jiang, H., Mahan, Thomas E., Ma, S., Maloney, Susan E., Wozniak, David F., Diamond, Marc I., and Holtzman, David M. (2013). Anti-Tau Antibodies that Block Tau Aggregate Seeding In Vitro Markedly Decrease Pathology and Improve Cognition In Vivo. Neuron *80*, 402-414.

Yen, S., Easson, C., Nacharaju, P., Hutton, M., and Yen, S.-H. (1999). FTDP-17 tau mutations decrease the susceptibility of tau to calpain I digestion. FEBS Letters *461*, 91-95.

Yuan, A., Kumar, A., Peterhoff, C., Duff, K., and Nixon, R.A. (2008). Axonal Transport Rates In Vivo Are Unaffected by Tau Deletion or Overexpression in Mice. The Journal of Neuroscience *28*, 1682-1687.

Yuzwa, S.A., Macauley, M.S., Heinonen, J.E., Shan, X., Dennis, R.J., He, Y., Whitworth, G.E., Stubbs, K.A., McEachern, E.J., Davies, G.J.*, et al.* (2008). A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo. Nature Chemical Biology *4*, 483-490.

Yuzwa, S.A., Shan, X., Macauley, M.S., Clark, T., Skorobogatko, Y., Vosseller, K., and Vocadlo, D.J. (2012). Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. Nature Chemical Biology *8*, 393-399.

Zhong, Q., Congdon, E.E., Nagaraja, H.N., and Kuret, J. (2012). Tau Isoform Composition Influences Rate and Extent of Filament Formation. Journal of Biological Chemistry *287*, 20711-20719.

Zhukareva, V., Mann, D., Pickering-Brown, S., Uryu, K., Shuck, T., Shah, K., Grossman, M., Miller, B.L., Hulette, C.M., Feinstein, S.C.*, et al.* (2002). Sporadic Pick's disease: A tauopathy characterized by a spectrum of pathological τ isoforms in gray and white matter. Annals of Neurology *51*, 730-739.

Chapter 2: Genome-wide scans for PSP risk alleles

Introduction

Tau pathology is a prominent hallmark of neurodegenerative diseases, including Alzheimer's Disease (AD) and Frontotemporal Dementia (FTD). PSP is a relatively pure tauopathy associated with parkinsonism - dementia, characterized by pathological tau aggregation and a clinical syndrome of postural instability, falls, and supranuclear opthalmoplegia (Respondek et al., 2013). It shares symptomatic and neuropathologic overlap with a large group of diseases, that are collectivity known as "tauopathies" due to characteristic tau deposits; however, compared to these diseases, PSP appears to be more clinically, neuropathologically, and genetically homogenous (Chen et al., 2015; Josephs et al., 2006; Williams et al., 2005). Notably, the clinical syndrome has high correlation with the neuropathology (Osaki et al., 2004). The major known genetic risk factor is an extended H1 haplotype on chromosome 17q21.31, which includes *MAPT* (the gene encoding the tau protein), and is homozygous in almost all PSP patients (Chen et al., 2015). Other risk factors identified include genome-wide significant associations at loci near *MAPT*, *MOBP*, *STX6*, and *EIF2AK3*, suggesting a strong contribution of common variation in its genetic architecture (Hoglinger et al., 2011). We reasoned that the inclusion of additional cases and controls could increase the statistical power for genome-wide association, potentially yielding novel loci that could provide insight into the molecular mechanisms of PSP and other more common tauopathies.

GWAS Cohort

We analyzed subjects from three GWAS cohorts, including 1) a multi-center cohort (Boxer et al., 2014; Chen et al., 2015) in whom we performed genotyping using the Illumina HumanOmni2.5-8 BeadChip and the Illumina HumanCore BeadChip ("UCLA"); 2) a cohort of autopsy-proven cases from previously published (Hoglinger et al., 2011) GWAS ("Hoglinger"); and 3) patients from centers in France, Germany, and the United Kingdom as part of the Neuroprotection and

Natural History in Parkinson Plus Syndromes (NNIPPS) study, a double-blind randomized placebo-controlled clinical trial of riluzole (Bensimon et al., 2009), genotyped with the Illumina HumanOmni2.5-8 BeadChip ("NNIPPS"). A more detailed description of the cohorts is provided in Supplementary Table A-1.

To increase statistical power, we combined each cohort with platform-matched, out-of-sample controls from dbGAP (Supplementary Table A-1). Stringent quality control – excluding SNPs that had low genotype call rates (< 0.95) or did not follow Hardy-Weinberg equilibrium, and excluding subjects with low sample call rate (< 0.95), non-European ancestry, incompatible sex, cryptic relatedness, or duplication across cohorts (Supplementary Figure A-1) – was applied to each cohort (including platform-matched controls). We then imputed variants implementing IMPUTE2 (Howie et al., 2009) using the 1000 Genomes Phase 3 Reference Panel to estimate genotypes at more than 77,000,000 SNPs. Imputed variants with imputation quality scores (r^2 < 0.9) or low minor allele frequency (< 0.01) were filtered, and genotypes across all cohorts were combined in a joint analysis. In total, we examined 6,419,662 SNPs in 1,646 PSP cases and 10,662 controls. We estimated that this cohort had 90% power to detect association of a variant with allele frequency of 0.5 and relative risk of 1.3. For a cohort of the sample size of that in a previous PSP GWAS from Hoglinger et al., the power to detect such an association was only 33%. In the primary analysis, we assessed the genome-wide association between the genotype at each SNP and case-control status using a linear mixed model to correct for population stratification. The genomic inflation factor λ for the joint analysis was 1.05; for the UCLA-Omni2.5, UCLA-HumanCore, NNIPPS, and Hoglinger cohorts, λ was 1.03, 1.02, 1.11, and 1.11, respectively (Figure 2-1, Supplementary Figure A-2, Supplementary Figure A-3). We considered the joint inflation factor to be acceptable in the setting of a relatively large joint analysis sample size (Yang et al., 2011b). Scaled for sample size, the adjusted genomic inflation factor λ_{1000} was 1.02.

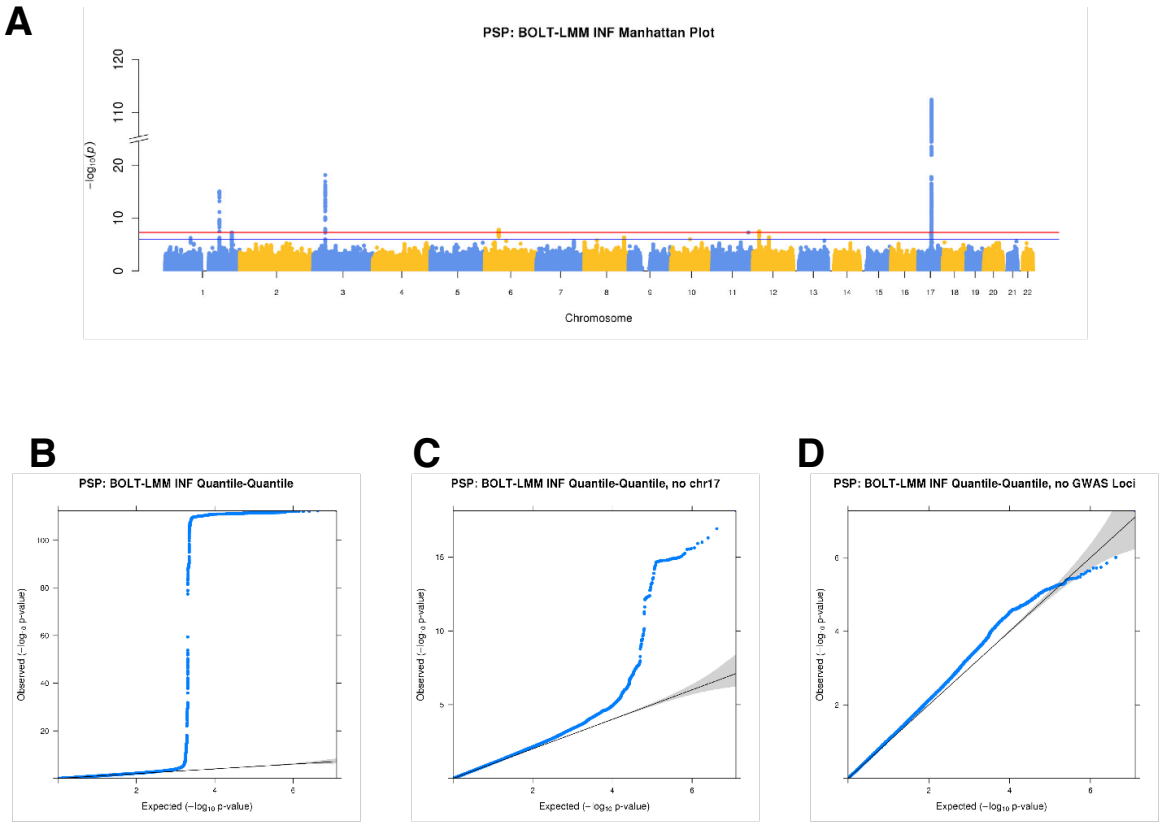


Figure 2-1: Genome-wide significant SNP associations in the joint analysis.
A) *P*-values of SNPs measured and imputed in the study. The horizontal red line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$). B) Quantile-quantile (QQ) plot showing the observed versus expected *P*-values for all SNPs typed in the study. C) QQ plot for SNPs, excluding chromosome 17. D) QQ plot for SNPs, excluding suggestive and genome-wide significant loci.

Copy-number variation analysis

A total of 4,866 CNVs in 281 PSP patients and 1,084 controls passed quality control. Of these, 2,769 CNVs including 1,205 deletions and 1,564 duplications were present at less than 1% frequency (and therefore considered rare). We did not detect an increased burden of rare CNVs in PSP patients after adjusting for multiple comparisons, even when stratifying by CNV size (Supplementary Table A-2). A trend towards enrichment of 200 kb – 500 kb duplications was detected in PSP (mean of 0.34 segments in PSP patients versus mean of 0.26 segments in controls, unadjusted p = 0.02). To identify rare CNVs that might cause PSP, recurrent genic CNVs (found in two or more PSP cases but not in WTCCC controls) were identified. We found

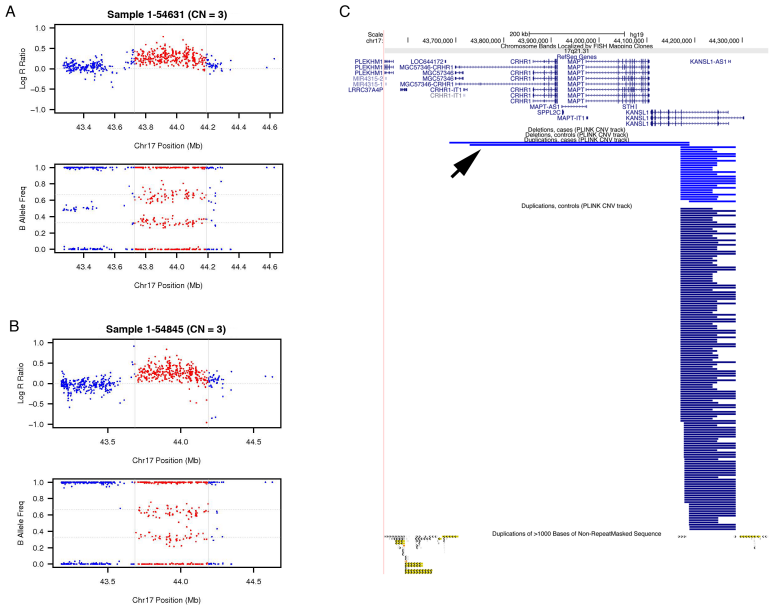


Figure 2-2: Recurrent rare CNVs within the 17q21.31 region.
A CNV was identified in a 40-year old man with PSP within 17q21.31 based on signal intensity (A). A second CNV with similar breakpoints in chr17q21.31 was found in a 62-year old woman with PSP (B). These CNVs overlapped *MAPT* and neighboring genes, as demonstrated on the UCSC Genome Browser Track (C; arrowhead). Segmental duplications are also shown.

four recurrent duplications, located within 2q37.1, 4q31.21, 16p12.2, and 17q21.31 (Supplementary Figure A-4). Each recurrent CNV was detected in two PSP patients.

We focused on the two duplications in 17q21.31 (Figure 2-2). We identified a 40-year-old European French male patient with autopsy-confirmed PSP and an unusually early age of onset of 37, who carried a 460 kb duplication, copy number = 3, at chr17:43,728,377-44,189,068 (hg19). A similar duplication (copy number = 3) was found in a 62-year-old European French woman with autopsy-confirmed PSP diagnosed at age 57, affecting 503 kb at chr17:43,685,926-44,189,068. These duplications spanned the entirety of the *MAPT* gene, as well as the last seven exons of *CRHR1* and the first six exons of *KANSL1*.

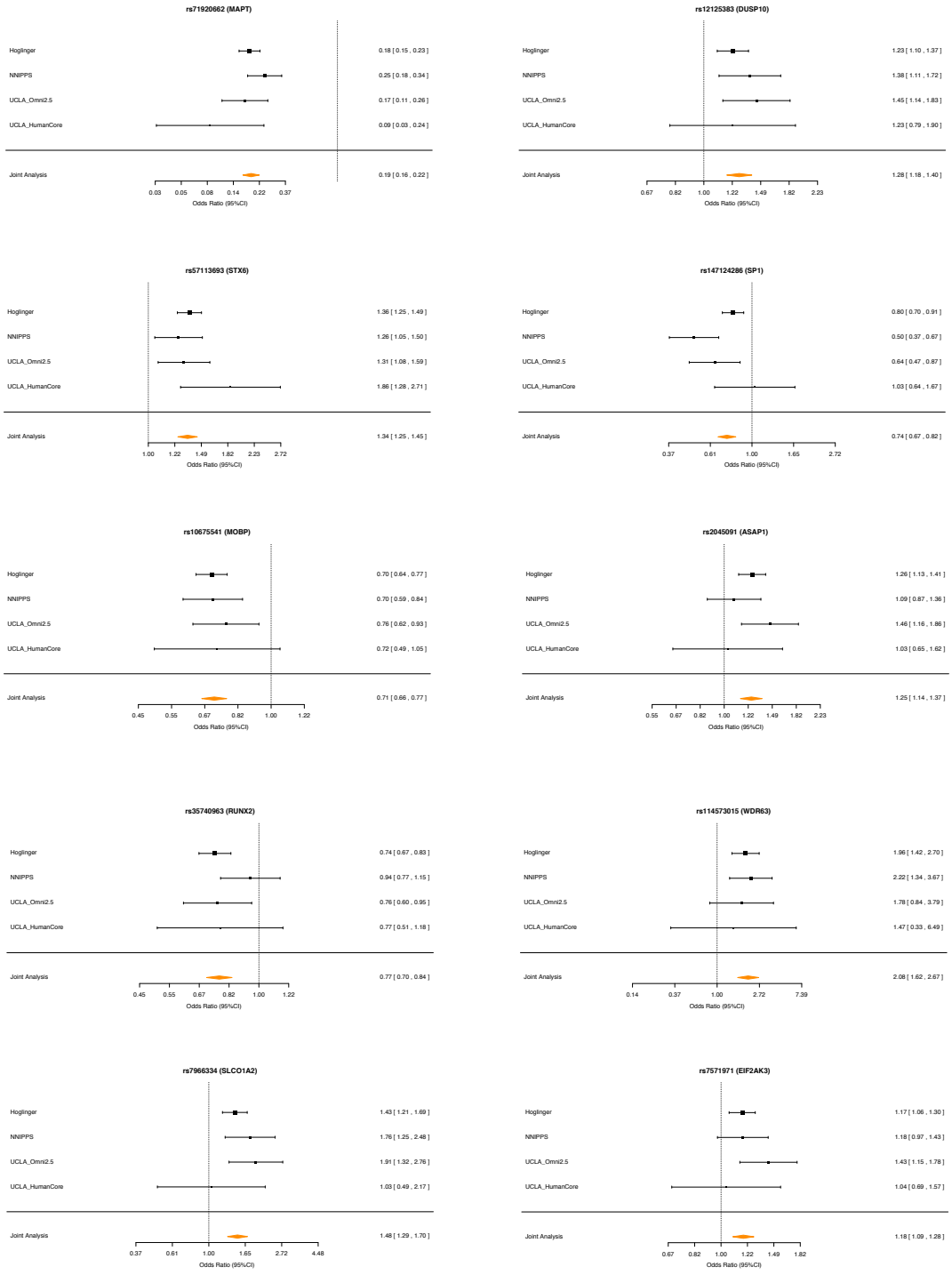


Figure 2-3: Meta-analysis of association for significant and suggestive SNPs. A total of eight genome-wide significant loci were identified, with representative SNPs (up to down, left to right): A) rs71920662 in 17q21.31, near MAPT; B) rs57113693 in 1q25.3, near STX6; C) rs10675541 in 3p22.1, near *MOBP*; D) rs35740963 in 6p21.1, near *RUNX2*; E) rs7966334 in 12p12.1, near *SLCO1A2*; F) rs12125383 in 1q41, near *DUSP10* in an intergenic region; G) rs147124286 in 12q13.13, near *SP1*; H) rs2045091 in 8q24.21, near *ASAP1*; and I) rs114573015 in 1p22.3, near *WDR63*. I) Additionally, a previously reported GWAS SNP rs7571971 in 2p11.2, near *EIF2AK3*, was not identified as genome-wide significant in the joint analysis.

Genome-wide SNP association meta-analysis

The results of the joint analysis genome-wide association are shown in Figure 2-1 and Supplementary Figure A-5. SNPs at 5 loci, in cytobands 17q21.31 (in an extended haplotype containing *MAPT*, lead SNP rs71920662, odds ratio OR = 0.19, $p = 3.9 \times 10^{-113}$), 3p22.1 (within *MOBP*, rs10675541, OR = 0.71, $p = 7.2 \times 10^{-19}$), 1q25.3 (within *STX6*, rs57113693, OR = 1.3, $p = 8.7 \times 10^{-16}$), 6p21.1 (within *RUNX2*, rs35740963, OR = 0.77, $p = 1.8 \times 10^{-8}$), and 12p12.1 (within *SLCO1A2*, rs7966334, OR = 1.5, $p = 3.2 \times 10^{-8}$), reached genome-wide significance ($p < 5 \times 10^{-8}$) (Supplementary Figure A-6). An additional SNP reported in a previous GWAS (Hoglinger et al., 2011), rs7571971, was also analyzed. Although this SNP did not reach genome-wide significance in the joint analysis (OR = 1.18, $p = 2.7 \times 10^{-5}$), the direction of the association was consistent with the previous association in each cohort. In order to decrease the likelihood that the results were influenced by population stratification, we assessed association at the loci in each of the study cohorts (Figure 2-3). Associations at the lead SNPs in each of the regions were consistent across the three most well-powered study cohorts (Hoglinger, NNIPPS, and UCLA Omni2.5) while in general, the HumanCore subset of the UCLA cohort was underpowered to detect association. An additional 4 loci demonstrated suggestive association ($1 \times 10^{-6} < P < 5 \times 10^{-8}$), in 1q41 (intergenic, near *DUSP10*, rs12125383, OR = 1.28, $p = 5.3 \times 10^{-8}$), 12q13.13 (within *SP1*, rs147124286, OR = 0.74, $p = 4.1 \times 10^{-7}$), 8q24.21 (within *ASAP1*, rs2045091, OR = 1.25, $p = 4.7 \times 10^{-7}$), and 1p22.3 (near *WDR63* and *MIR4423*, rs114573015, OR = 2.1, $p = 5.9 \times 10^{-7}$). Overall, the genome-wide significant loci explained a combined 5.9% of the variance in heritable liability of PSP (Supplementary Table A-3). The locus tagging the chr17q21 haplotype surrounding MAPT contributed the majority (5.0%), while new loci contributed an additional 0.2% of the total liability. Using a polygenic model implemented in GCTA (ref), the entire set of genotyped SNPs explains 9.4±0.8% (estimate±standard error) of the variance on the liability scale, suggesting that many loci are yet to be found.

The association between PSP and the chr17q21 haplotype (H1/H2) has been widely characterized, but independent SNPs in the chr17q21 region may also contribute to disease susceptibility. To test this, we performed linear regression, taking haplotype as a covariate. Additionally, we identified subjects that were homozygous for the risk allele (H1/H1), and performed association in this subset of patients. Both approaches identified similar independent associations from the H1 haplotype in the 17q21.31 region, with the most significant SNPs at rs8078967 (P = 1.9 x 10^{-14}) and rs9904290 (P = 8.9 x 10^{-12}) in the haplotype-regressed and H1/H1 only datasets, respectively (Supplementary Figure A-6). These SNPs did not appear to be in strong linkage disequilibrium with a previously reported SNP association, rs242557(Hoglinger et al., 2011) (r^2 = 0.008 and 0.007 in the 1000 Genomes Project data – EUR super-population, respectively) that was filtered from this dataset in variant QC; however, they were highly correlated with each other (r^2 = 0.996). Additionally, both variants and rs242557 are within the first intron of the *MAPT* gene.

Fine-mapping of associated loci

To further understand how variation at each of the loci contributes to disease risk, we assessed the functional consequences of significant SNPs. We first identified a set of potential causal SNPs using the CAVIAR method, which identifies a "credible set" of SNPs that encompasses those likely to be causal (Hormozdiari et al., 2014). The 17q21.31 locus was excluded from the analysis because of its unusual, long-range linkage disequlibrium pattern. In some loci, potentially causal coding variants were identified (in genome-wide significant loci, at 6p21.1, in *RUNX2*, and at 12p12.1, in *SLCO1A2*; and in suggestive loci, at 8q24.21, in *ASAP1*, and at 12q13.13, in *AMHR2*). Other SNPs in the credible set fell within regulatory regions; we identified the gene associated with each SNP using data from Hi-C experiments, mapping chromosome conformation patterns on a genome-wide scale from four human cell types (IMR-90 fetal lung

fibroblasts, embryonic stem cells, fetal brain, and fetal brain germinal zone). Each potential regulatory SNP in the credible set was then associated with genes in close proximity by chromosomal conformation, yielding potential downstream causal genes.

To supplement the mapping information from HiC, we also identified the functional consequences of GWAS hits, we used the TWAS method to predict genes that may be affected by risk alleles (Gusev et al., 2016). TWAS estimates gene expression values using paired reference transcriptome/genotyping datasets (e.g., for expression quantitative trait loci - eQTL studies) and genotype information from summary statistics, and predicts differential expression between cases and controls. Using reference data from the GTEx Consortium, TWAS predicted the effect of gene expression from the risk haplotypes in multiple tissues. At a threshold of $P < 1x10^{-5}$, we identified a number of genes that were called as differentially expressed. As expected due to the length and lack of recombination in the region, most of these genes (17) clustered around the chromosome 17 haplotype. Notably, *MAPT* (within the associated 17q21.31 locus) was among the genes predicted to be differentially expressed, as well as *STX6* (within the associated 1q24 locus), *SP1* (within the suggestive 12q13.13 locus), *SKIV2L* (within 6p21.33, nearby the associated 6p21.1 locus), and *RPSA* (within the associated 3p22.1 locus). Other genes that were pinpointed outside of association regions were *CEP57* (in 11q21) and *RPS6KL1* (in 14q24.3)

Genetic overlap of PSP and other neurodegenerative diseases

The strong neuropathological overlap of PSP with other tauopathies suggests that genetic overlap may exist. Using the LDSC software (Bulik-Sullivan et al., 2015), we assessed genetic overlap of PSP with other neurodegenerative diseases, including AD, behavioral variant FTD (bvFTD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS), by using GWAS summary statistics. As controls, we included summary statistics from GWAS for heritable, non-

neurodegenerative diseases of brain (schizophrenia and bipolar disorder), a quantitative trait (height), and a non-brain disease (type 2 diabetes) (for further details, refer to the Supplementary Methods). Each of these traits was shown to be heritable. Statistically significant genetic correlations were identified for PD (P = 9.7x10^{-5}) and ALS (P = 1.8x10^{-3}), but not for non-neurodegenerative disease control GWAS (Figure 2-4).

Discussion

As a prototypical tauopathy, insight into PSP susceptibility alleles can help to illuminate the downstream molecular effects of tau pathology, which is a major component of many common neurodegenerative diseases. Altogether, from a joint analysis of three disease cohorts, we have identified 2 novel genome-wide significant susceptibility loci in PSP and replicated 3 previously reported loci. Of the loci identified in this study, three (within *MAPT*, *MOBP*, and *STX6*) were reported significant in a previous GWAS (Hoglinger et al., 2011). An additional locus near the *EIF2AK3* gene (encoding PERK, a key component of the unfolded protein response) was also previously identified; however, the reported SNP did not reach genome-wide significance in this joint analysis or in the new "Hoglinger" cohort (using different controls).

We also identified 2 novel genome-wide significant susceptibility loci at 6p21.1 and 12p12.1 (near *RUNX2* and *SLCO1A2*, respectively). At 6p21.1, we identified a lead SNP as well as several coding SNPs in *RUNX2* in the credible set within *RUNX2*, a gene thought to be a transcriptional factor involved in regulation of osteoblastic differentiation (Komori, 2009). While seemingly unrelated to PSP, a curious number of neurodegeneration-related genes are also involved in bone diseases (e.g. *TREM2*, which has been linked to AD and Nasu-Hakola disease (Guerreiro et al., 2013; Jonsson et al., 2013), and *VCP*, linked to amyotrophic lateral sclerosis and Paget's disease of bone (Johnson et al., 2010; Watts et al., 2004)). At 12p12.1, we identified a lead SNP and credible set coding SNPs within *SLCO1A2*, a transporter present

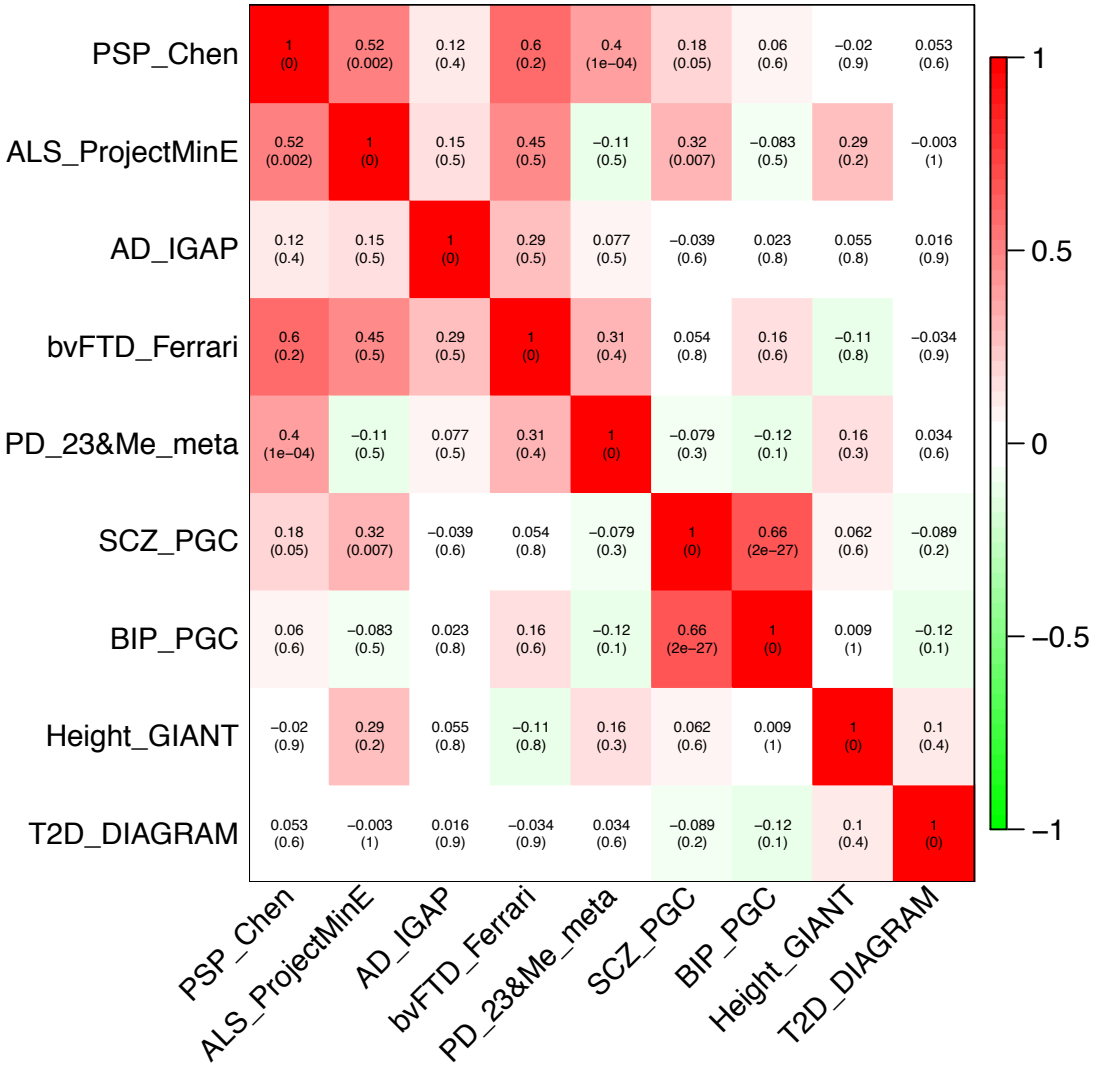


Figure 2-4: Coheritability based upon LD Score Regression for neurodegenerative diseases. Each cell contains the genetic correlation and p-value (in parentheses).

(among other places) at the blood-brain barrier, where it regulates solute trafficking (Urquhart and Kim, 2009). An additional four loci (near the genes *DUSP10*, *SP1*, *ASAP1*, and *WDR63*/*MIR4423*) were suggestive of association, but did not reach genome-wide significance. While this study raises the possibility of involvement of these genes in PSP pathogenesis, further fine-mapping and functional studies are needed to confirm their possible roles.

Our results also implicate possible alternative causal genes in previously reported genome-wide significant loci. At 3p22.1, the gene closest to the GWAS lead SNP was reported as *MOBP*. Using Hi-C, we have identified chromatin interactions with *MYRIP* and *EIF1B* that could also explain this association. Similarly, at 1q25.3, the gene closest to the GWAS lead SNP was *STX6*; by Hi-C, we have also identified *XPR1* as a possible candidate gene. Interestingly, our group has previously demonstrated *XPR1* mutations in primary familial brain calcification (Legati et al., 2015), though any mechanistic overlap with PSP is unclear. Analysis of eQTL datasets (in GTEx) suggests that *RPSA* at 3p22.1 and *SKIV2L* near 6p21.1 may also be the causal gene, but the tissue-relevant datasets were relatively underpowered.

Aside from identifying additional associated loci and highlighting potential PSP susceptibility genes, we analyzed the polygenic overlap between neurodegenerative diseases, identifying shared heritability with PD and ALS. Curiously, these diseases do not have predominant tau neuropathology, as PSP and other tauopathies do. Typically, PD is associated with aggregation of α-synuclein, and ALS with aggregation of TDP43 and other proteins, while tau pathology is prominent in AD. However, there are known shared genetic risk factors among these diseases. The 17q21.31 haplotype is highly associated with PD, in the same direction as in PSP (Nalls et al., 2016), and SNPs near the *MOBP* gene have been recently associated with ALS (van Rheenen et al., 2016). Our results indicate the existence of common neurodegenerative disease pathways even across traditional protein aggregate-based subdivisions, and could potentially lead to effective treatment strategies.

We also performed a scan for copy-number variants in one of the cohorts. We found two instances of CNV of *MAPT* in patients with PSP. Taken together with a copious body of evidence connecting *MAPT* with PSP genetically and neuropathologically, this finding suggests that duplication of *MAPT* is a significant, though relatively uncommon cause of PSP.

Duplications spanning the *MAPT* gene have been previously reported in twelve patients, including in children with mild learning difficulties, developmental delay, and variable dysmorphic features (Gregor et al., 2012); a patient with a presumed familial frontotemporal dementia phenotype (Rovelet-Lecrux et al., 2010) and three siblings with early-onset Alzheimer's disease (Hooli et al., 2014). Other studies have not detected *MAPT* duplications in their neurodegenerative disease cohorts (Llado et al., 2007; Skoglund et al., 2009), highlighting the duplication's causal role in PSP.

The localisation of the observed duplications also narrows the search space and provides additional mechanistic insight compared to GWAS SNPs alone. Although the CNVs span several genes, the region is much smaller than the approximated 1Mb *MAPT* haplotype on which previously identified duplications resulted in developmental anomalies. Furthermore, the putative effect on gene expression is much clearer, while the 17q21.31 haplotype does not seem to affect overall expression and probably alters splicing of the *MAPT* exon 3 (Trabzuni et al., 2012). Overall, the association of *MAPT* duplication with PSP for the first time provides direct genetic evidence that specific overexpression of tau can lead to PSP. This is consistent with data from models showing that human wild-type tau over-expression is sufficient to cause pathological tau deposition and neurodegeneration, even in the absence of mutations (Jackson et al., 2002).

In addition to the two patients harbouring duplications of the entire tau locus, we found several additional genic regions with rare, recurrent CNVs in PSP patients, including duplications on 2q37.1 (spanning *ALPP, ECEL1P2*, and *ALPPL2*), 4q31.21 (spanning *RNF150*, and *ZNF330*), and 16p12.2 (spanning *METTL9, IGSF6,* and *OTOA*). However, given our sample size and the rarity of these events, the pathogenicity of these variants is difficult to establish without prior evidence of involvement in PSP. Our finding of rare CNVs within *MAPT* in PSP strongly

suggests that CNVs can cause PSP in a highly penetrant fashion, and future work building on our results may yet identify additional PSP genetic contributors within these regions.

Here, we have increased the number of significant genetic risk locus for PSP, an important advance for understanding its pathophysiology. The power of this study to identify novel loci at genome wide significance and a large unexplained heritability suggests that PSP may be highly amenable to genetic association studies in larger sample cohorts using next generation sequencing. Overall, by establishing the genetic correlations of PSP with PD and ALS and identifying novel genome-wide significant and suggestive associations, we shed insight into the mechanisms of neurodegenerative disease.

Methods

Cohort. Three cohorts of primarily European ancestry were included in the study – "UCLA", a combination of PSP patients and controls from the UCSF Memory and Aging Center (Chen et al., 2015; Li et al., 2014) and the Allon Therapeutics Davunetide trial (Boxer et al., 2014); "NNIPPS", a group of PSP patients from the Neuroprotection and Natural History in Parkinson Plus Syndromes (NNIPPS) trial (Bensimon et al., 2009); and "Hoglinger", PSP patients from a previously published GWAS (Hoglinger et al., 2011). The UCLA cohort was divided into two, because of differences in genotyping platform: "UCLA Omni 2.5" and "UCLA HumanCore". Further details are available in the Supplementary Methods.

Genotyping. Genotyping in the UCLA study cohort was performed as a prelude to whole-genome sequencing, and was performed by Illumina (using the Illumina HumanOmni 2.5 Array) and the New York Genome Center (using the Illumina HumanCore Array).

Public datasets. Genotypes from the Hoglinger et al. GWAS (2011) (cases only – no controls) were obtained from the NIAGADS database. Out-of-sample controls were obtained from dbGAP Authorized Access to match each genotyping platform. For the Illumina HumanQuad 660W Array (Hoglinger et al. study), we used phs000103.v1.p1 "Genome-Wide Association Studies of Prematurity and Its Complications", phs000289.v1.p1 "National Human Genome Research Institute (NHGRI) GENEVA Genome-Wide Association Study of Venous Thrombosis", phs000188.v1.p1 "Vanderbilt Genome-Electronic Records (VGER) Project: QRS Duration", phs000203.v1.p1 "A Genome-Wide Association Study of Peripheral Arterial Disease", phs000237.v1.p1 "Northwestern NUgene Project: Type 2 Diabetes", phs000243.v1.p1 "Group Health/UW Aging and Dementia eMERGE study", and phs000170.v1.p1 "A Genome-Wide Association Study on Cataract and HDL in the Personalized Medicine Research Project Cohort". For the Illumina HumanOmni 2.5 Array (UCLA – this study, and NNIPPS study), we used phs000371.v1.p1 "Genetic Modifiers of Huntington's Disease", phs000429.v1.p1 "NEI Age-Related Eye Disease Study (AREDS) - Genetic Variation in Refractive Error Substudy", and phs000421.v1.p1 "A Genome-Wide Association Study of Fuchs' Endothelial Corneal Dystrophy (FECD)". For the Illumina HumanCore Array (UCLA – this study), we used the WTCCC2 cohort, which was typed on the related Illumina OmniExpress Array. Subjects with an ascertained phenotype (e.g., disease) were removed. More detailed information regarding these datasets is available in Supplementary Table A-1.

Data preprocessing. Genotypes for all datasets were converted to the forward strand, and converted into coordinates based on the hg19 reference sequence using UCSC liftOver (Rosenbloom et al., 2015). The genotypes were then merged and pre-processed according to platform. Determination of cryptic relatedness (pairwise proportion IBD, PI-HAT > 0.2), sample missingness (> 0.05), genotype missingness (> 0.05), Hardy-Weinberg equilibrium p-value (< 10^5), and sex-matching was performed in PLINK v1.90b3.28 (Chang et al., 2015) and used to

quality-control samples using standard parameters (Evangelou and Ioannidis, 2013). Ancestry was predicted by multidimensional scaling based on raw Hamming distances, implemented in PLINK. Only samples of presumed European ancestry that clustered with known Europeans from the HapMap3 cohort (The International HapMap 3 Consortium, 2010) were included. Preprocessing steps are further elaborated in Supplementary Figure A-1.

Imputation. Imputation was performed separately for each genotyping platform using the IMPUTE v2.3.2 (Howie et al., 2009) algorithm. Prephasing of chromosomes using the Segmented HAPlotype Estimation & Imputation Tool (SHAPEIT) v2.r837 was performed as previously described (Delaneau et al., 2014; O'Connell et al., 2014). IMPUTE2 was run on the prephased haplotypes using the 1000 Genomes Project Phase 3 reference in non-overlapping 5 megabase chunks with a 250 kilobase buffer and an effective population size of 20,000. Imputed variants with an imputation genotype probability < 0.9, missingness > 0.05, or minor allele frequency < 0.01 were removed, and genotypes across platforms were merged. Cryptic relatedness across cohorts was assessed, and related/duplicated samples were removed.

Copy number variation analysis. CNVs were identified from genotyping data of patients with PSP and controls. Initially, sample reclustering was performed on each sample using Illumina GenomeStudio, for each array batch. Samples with low call rates (< 98%) were excluded. To compare CNVs between PSP cases and controls from the WTCCC_1958, only SNPs in common between the two array platforms were used for subsequent analysis (689,077 SNPs in total). PennCNV (Wang et al., 2007) was used to call CNVs, using custom PFB and GC model files and with genomic wave adjustment. Adjacent CNV calls were merged if their separation spanned < 20% their combined length. CNVs overlapping (> 50%) immunoglobulin, telomeric, and centromeric regions; called on < 10 array SNPs; spanning < 50,000 base pairs; and having confidence score < 10 were filtered. Subjects with LRR standard deviation > 0.285; BAF Drift <

0.01; Waviness Factor < 0.05; total called CNVs > 100; and maximum combined CNV size < 10,000,000 base pairs were removed from further analysis. CNVs highlighted by downstream analyses were validated by manual examination of the signal intensity.

Rare CNVs (frequency < 1%) were considered for downstream analysis. CNV burden and association testing was performed using PLINK v1.07 and the Bedtools package (Quinlan and Hall, 2010). Empirical p-values were calculated using the maxT test with 50,000 permutations, with statistical significance defined at corrected $p < 0.05$.

Association. Association was performed using a linear mixed model to correct for population structure, using BOLT-LMM (Loh et al., 2015). The genotyping platform was used as a categorical covariate. The standard infinitesimal model p-values were chosen for downstream anaylsis. Odds ratios were calculated as exp(beta). Because some of the individual cohort sizes violate the large sample size assumptions of BOLT-LMM, odds ratios for association (for individual cohorts) were computed using a logistic regression model in PLINK, using the first 5 eigenvectors, derived from Principal Components Analysis (PCA), as covariates. Power calculations were performed using the Genetic Power Calculator (Purcell et al., 2003), assuming a variant with risk allele frequency of 0.5 and relative risk of 1.3 in an additive genetic model, a disease with a prevalence of 10 in 100,000, and a p-value threshold of $5x10^{-8}$, using a genotypic, 2 df case-control test. QQ and Manhattan plots were constructed using the R package "qqman" (Turner, 2014). Forest plots were constructed using the R package "metafor" (Viechtbauer, 2010). The genomic inflation factor λ was computed with PLINK. Correction of the genomic inflation factor to an equivalent sample size of 1000 cases and 1000 controls was performed as previously described (Freedman et al., 2004). To control for the extended haplotype on chr17q21 and to identify independent association signals, we performed

association as before, but including the haplotype (tagged by the SNP rs1560310) (Coppola et al., 2012) as a covariate.

Proportion variance in liability explained. The explained variance in liability at each of the genome-wide significant loci was calculated according to the method of So et al., which requires the allele frequency of the risk allele, the relative risk of the heterozygous genotype, the relative risk of the homozygous risk genotype, and the prevalence of the disease in the population (So et al., 2011). The allele frequencies were calculated from the control population of the joint genotyping cohort. Relative risks were approximated with the corresponding odds ratios, which convergence to relative risk when the prevalence of disease is rare. Genotypic odds ratios were estimated by assuming an additive model. The prevalence of PSP was estimated at 6.5 per 100,000 in accordance with prior epidemiological studies (Nath et al., 2001; Schrag et al., 1999). The genome-wide polygenic variance in liability explained was calculated using GCTA v1.24.7 (Yang et al., 2011a). The genetic relationship matrix was calculated chromosome-by-chromosome and then re-combined. The first 5 principal components were calculated and used as covariates for restricted maximum likelihood (REML) analysis.

Prediction of gene expression differences associated with PSP-associated SNPs. Genetic associations with PSP may be due to genetic control of gene expression. We used TWAS to predict differential gene expression in PSP from the joint analysis summary statistics, integrating paired genotyping and gene expression data from the GTEx Consortium (Gusev et al., 2016). Correcting for approximately 5,000 effective independent tests per brain region (taking into account 5,483 genes with significantly heritable weights and the interdependence of gene expression, particularly across tissues), the significance threshold was set at $P < 1 \times 10^{-5}$.

Credible set of causal variants at PSP GWAS loci. A credible set (potential causal variants) was identified at each of total of seven genome-wide significant loci identified in this study using the CAusal Variants Identification in Associated Regions (CAVIAR) software package (Hormozdiari et al., 2014). Because of the extended linkage disequilibrium patterns in the chromosome 17q21.31 haplotype region, causal variants were not identified at this associated locus. Within each of the selected loci, the SNP with the minimum joint association p-value was chosen as the index SNP, and variants with p-value < 10^{-5} and in LD (r^2 > 0.6) with the index SNP were input into CAVIAR. The CAVIAR-identified credible set contains potential causal variants (with a confidence level of 95% under the statistical model) that could explain the association at each locus.

Identification of genes linked to credible SNPs with chromatin interaction data. Genetic variation can result in changes to the coding sequence of a gene (e.g., nonsense and missense variants) or can regulate the gene's expression (e.g., by affecting transcription factor binding in promoter or enhancer regions). We first identified credible SNPs as "functional" (stopgain variant, frameshift variant, splice donor variant, NMD transcript variant, or missense variant). Of the remaining credible SNPs, we identified those in the promoter region of a gene, defined as the range 2 kb upstream to 1 kb downstream relative to the transcription start site (TSS). Finally, the remaining credible SNPs were considered possible regulatory variants and tested for short- or long- range interaction with other regions of chromatin to identify potential downstream target genes. The interactions were determined by Hi-C experiments in IMR90 and embryonic stem cells from public data (Dixon et al., 2015; Dixon et al., 2012), and fetal brain germinal zone (ventricular and subventricular zone) and cortical plate (intermediate zone and marginal zone) from our group (Won et al., 2016).

Genetic correlation with neurodegenerative diseases. Genetic correlation was assessed from GWAS summary statistics using the Linkage Disequilibrium Score Regression method (LDSC) (Bulik-Sullivan et al., 2015). Summary statistics were filtered by only considering SNPs that overlap with the HapMap3 reference panel. Refer to the Supplementary Methods for further details.

References

Bensimon, G., Ludolph, A., Agid, Y., Vidailhet, M., Payan, C., and Leigh, P.N. (2009). Riluzole treatment, survival and diagnostic criteria in Parkinson plus disorders: The NNIPPS Study. Brain *132*, 156-171.

Boxer, A.L., Lang, A.E., Grossman, M., Knopman, D.S., Miller, B.L., Schneider, L.S., Doody, R.S., Lees, A., Golbe, L.I., Williams, D.R.*, et al.* (2014). Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. The Lancet Neurology *13*, 676-685.

Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics, C., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics *47*, 291-295.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 1-16.

Chen, J.A., Wang, Q., Davis-Turak, J., and et al. (2015). A multiancestral genome-wide exome array study of alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy. JAMA Neurology *72*, 414-422.

Coppola, G., Chinnathambi, S., Lee, J.J., Dombroski, B.A., Baker, M.C., Soto-Ortolaza, A.I., Lee, S.E., Klein, E., Huang, A.Y., Sears, R.*, et al.* (2012). Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. Human Molecular Genetics *21*, 3500-3512.

Delaneau, O., Marchini, J., and The Genomes Project, C. (2014). Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nature Communications *5*.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W.*, et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331-336.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Evangelou, E., and Ioannidis, J.P.A. (2013). Meta-analysis methods for genome-wide association studies and beyond. Nature Reviews Genetics *14*, 379-389.

Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N.*, et al.* (2004). Assessing the impact of population stratification on genetic association studies. Nature Genetics *36*, 388-393.

Gregor, A., Krumbiegel, M., Kraus, C., Reis, A., and Zweier, C. (2012). De novo triplication of the MAPT gene from the recurrent 17q21.31 microdeletion region in a patient with moderate intellectual disability and various minor anomalies. Am J Med Genet A *158a*, 1765-1770.

Guerreiro , R., Wojtas , A., Bras , J., Carrasquillo , M., Rogaeva , E., Majounie , E., Cruchaga , C., Sassi , C., Kauwe , J.S.K., Younkin , S.*, et al.* (2013). TREM2 Variants in Alzheimer's Disease. New England Journal of Medicine *368*, 117-127.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A.*, et al.* (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics *48*, 245-252.

Hoglinger, G.U., Melhem, N.M., Dickson, D.W., Sleiman, P.M.A., Wang, L.-S., Klei, L., Rademakers, R., de Silva, R., Litvan, I., Riley, D.E.*, et al.* (2011). Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nature Genetics *43*, 699-705.

Hooli, B.V., Kovacs-Vajna, Z.M., Mullin, K., Blumenthal, M.A., Mattheisen, M., Zhang, C., Lange, C., Mohapatra, G., Bertram, L., and Tanzi, R.E. (2014). Rare autosomal copy number variations in early-onset familial Alzheimer's disease. Molecular Psychiatry *19*, 676-681.

Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B., and Eskin, E. (2014). Identifying Causal Variants at Loci with Multiple Signals of Association. Genetics.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLoS Genetics *5*, e1000529.

Jackson, G.R., Wiedau-Pazos, M., Sang, T.K., Wagle, N., Brown, C.A., Massachi, S., and Geschwind, D.H. (2002). Human wild-type tau interacts with wingless pathway components and produces neurofibrillary pathology in Drosophila. Neuron *34*, 509-519.

Johnson, J.O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V.M., Trojanowski, J.Q., Gibbs, J.R., Brunetti, M., Gronka, S., Wuu, J.*, et al.* (2010). Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS. Neuron *68*, 857-864.

Jonsson , T., Stefansson , H., Steinberg , S., Jonsdottir , I., Jonsson , P.V., Snaedal , J., Bjornsson , S., Huttenlocher , J., Levey , A.I., Lah , J.J.*, et al.* (2013). Variant of TREM2

Associated with the Risk of Alzheimer's Disease. New England Journal of Medicine *368*, 107-116.

Josephs, K.A., Petersen, R.C., Knopman, D.S., Boeve, B.F., Whitwell, J., Duffy, J.R., Parisi, J.E., and Dickson, D.W. (2006). Clinicopathologic analysis of frontotemporal and corticobasal degenerations and PSP. Neurology *66*, 41-48.

Komori, T. (2009). Regulation of bone development and extracellular matrix protein genes by RUNX2. Cell and Tissue Research *339*, 189-195.

Legati, A., Giovannini, D., Nicolas, G., Lopez-Sanchez, U., Quintans, B., Oliveira, J.R.M., Sears, R.L., Ramos, E.M., Spiteri, E., Sobrido, M.-J*., et al.* (2015). Mutations in XPR1 cause primary familial brain calcification associated with altered phosphate export. Nature Genetics *47*, 579-581.

Li, Y., Chen, J.A., Sears, R.L., Gao, F., Klein, E.D., Karydas, A., Geschwind, M.D., Rosen, H.J., Boxer, A.L., Guo, W*., et al.* (2014). An Epigenetic Signature in Peripheral Blood Associated with the Haplotype on 17q21.31, a Risk Factor for Neurodegenerative Tauopathy. PLoS Genetics *10*, e1004211.

Llado, A., Rodriguez-Santiago, B., Antonell, A., Sanchez-Valle, R., Molinuevo, J.L., Rene, R., and Perez-Jurado, L.A. (2007). MAPT gene duplications are not a cause of frontotemporal lobar degeneration. Neuroscience Letters *424*, 61-65.

Loh, P.-R., Tucker, G., Bulik-Sullivan, B.K., Vilhjalmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B*., et al.* (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nature Genetics *47*, 284-290.

Nalls, M.A., Keller, M.F., Hernandez, D.G., Chen, L., Stone, D.J., Singleton, A.B., and on behalf of the Parkinson's Progression Marker Initiative, i. (2016). Baseline genetic associations in the Parkinson's Progression Markers Initiative (PPMI). Movement Disorders *31*, 79-85.

Nath, U., Ben-Shlomo, Y., Thomson, R.G., Morris, H.R., Wood, N.W., Lees, A.J., and Burn, D.J. (2001). The prevalence of progressive supranuclear palsy (Steele–Richardson–Olszewski syndrome) in the UK. Brain *124*, 1438-1449.

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I*., et al.* (2014). A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. PLoS Genetics *10*, e1004234.

Osaki, Y., Ben-Shlomo, Y., Lees, A.J., Daniel, S.E., Colosimo, C., Wenning, G., and Quinn, N. (2004). Accuracy of clinical diagnosis of progressive supranuclear palsy. Movement Disorders *19*, 181-189.

Purcell, S., Cherny, S.S., and Sham, P.C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics *19*, 149-150.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Respondek, G., Roeber, S., Kretzschmar, H., Troakes, C., Al-Sarraj, S., Gelpi, E., Gaig, C., Chiu, W.Z., van Swieten, J.C., Oertel, W.H.*, et al.* (2013). Accuracy of the national institute for neurological disorders and stroke/society for progressive supranuclear palsy and neuroprotection and natural history in Parkinson plus syndromes criteria for the diagnosis of progressive supranuclear palsy. Movement Disorders *28*, 504-509.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M.*, et al.* (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Research *43*, D670-D681.

Rovelet-Lecrux, A., Hannequin, D., Guillin, O., Legallic, S., Jurici, S., Wallon, D., Frebourg, T., and Campion, D. (2010). Frontotemporal dementia phenotype associated with MAPT gene duplication. J Alzheimers Dis *21*, 897-902.

Schrag, A., Ben-Shlomo, Y., and Quinn, N.P. (1999). Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. The Lancet *354*, 1771-1775.

Skoglund, L., Ingvast, S., Matsui, T., Freeman, S.H., Frosch, M.P., Brundin, R., Giedraitis, V., Growdon, J.H., Hyman, B.T., Lannfelt, L.*, et al.* (2009). No evidence of PGRN or MAPT gene dosage alterations in a collection of patients with frontotemporal lobar degeneration. Dement Geriatr Cogn Disord *28*, 471-475.

So, H.-C., Gui, A.H.S., Cherny, S.S., and Sham, P.C. (2011). Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genetic Epidemiology *35*, 310-317.

The International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52-58.

Trabzuni, D., Wray, S., Vandrovcova, J., Ramasamy, A., Walker, R., Smith, C., Luk, C., Gibbs, J.R., Dillman, A., Hernandez, D.G.*, et al.* (2012). MAPT expression and splicing is differentially

regulated by brain region: relation to genotype and implication for tauopathies. Hum Mol Genet *21*, 4094-4103.

Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv.

Urquhart, B.L., and Kim, R.B. (2009). Blood−brain barrier transporters and response to CNS-active drugs. European Journal of Clinical Pharmacology *65*, 1063-1070.

van Rheenen, W., Shatunov, A., Dekker, A.M., McLaughlin, R.L., Diekstra, F.P., Pulit, S.L., van der Spek, R.A.A., Vosa, U., de Jong, S., Robinson, M.R.*, et al.* (2016). Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nature Genetics *48*, 1043-1048.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. 2010 *36*, 48.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H., and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Research *17*, 1665-1674.

Watts, G.D.J., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P., and Kimonis, V.E. (2004). Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. Nature Genetics *36*, 377-381.

Williams, D.R., de Silva, R., Paviour, D.C., Pittman, A., Watt, H.C., Kilford, L., Holton, J.L., Revesz, T., and Lees, A.J. (2005). Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. Brain *128*, 1247-1258.

Won, H., de la Torre-Ubieta, L., Stein, J.L., Parikshak, N.N., Huang, J., Opland, C.K., Gandal, M.J., Sutton, G.J., Hormozdiari, F., Lu, D.*, et al.* (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. Nature *538*, 523-527.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011a). GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics *88*, 76-82.

Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M.*, et al.* (2011b). Genomic inflation factors under polygenic inheritance. European Journal of Human Genetics *19*, 807-812.

Chapter 3: Genome-wide exome array study of Alzheimer's disease, frontotemporal dementia, and progressive supranuclear palsy

Introduction

Genetics studies have revealed a genetic contribution to susceptibility for common or sporadic forms of neurodegenerative disease, such as Alzheimer's disease (AD), frontotemporal dementia (FTD), and progressive supranuclear palsy (PSP – a syndrome characterized by oculomotor and gait abnormalities and usually associated with tau pathology). In Alzheimer's disease, early genetic mapping approaches have identified variants in genes such as *APP*, *PSEN1*, and *PSEN2* that cause familial, early-onset forms of AD (Lendon et al., 1997). *APOE* was also pinpointed as a late-onset AD susceptibility gene (Pericak-Vance et al., 1991). Genome-wide association studies (GWAS) targeted toward common variants in primarily European populations have identified many variants associated with Alzheimer's disease (AD), mostly clearly near *APOE* but also consistently near *ABCA7*, *BIN1*, *CLU*, *CR1*, *PICALM*, *SORL1*, and other genes (Hollingworth et al., 2011; Lambert et al., 2013; Naj et al., 2011). Next-generation sequencing approaches have also found rare variants with strong effect in genes such as *MAPT* (Coppola et al., 2012) and *TREM2* (Guerreiro et al., 2013).

In FTD, the most frequently observed mutations in familial cases occur in *C9ORF72* (a hexanucleotide repeat expansion), *GRN*, *MAPT* (associated with parkinsonism and sometimes manifesting as a PSP syndrome), *TARDBP*, and other genes (Rademakers et al., 2012). In sporadic cases, a haplotype variant on the long arm of chromosome 17 has been repeatedly associated with PSP (Baker et al., 1999; Hoglinger et al., 2011; Pittman et al., 2005). GWAS have also been performed for sporadic cases of FTD, identifying associated SNPs near *TMEM106B* in cases with TDP-43 pathology (Van Deerlin et al., 2010), *RAB38*/*CTSC* in behavioral variant FTD, and *BTNL2*/*HLA-DRA*/*HLA-DRB5* in clinically diagnosed FTD (Ferrari et al., 2014); and for PSP, identifying associated SNPs near *MAPT*, *MOBP*, *EIF2AK3*, and *STX6* (Hoglinger et al., 2011).

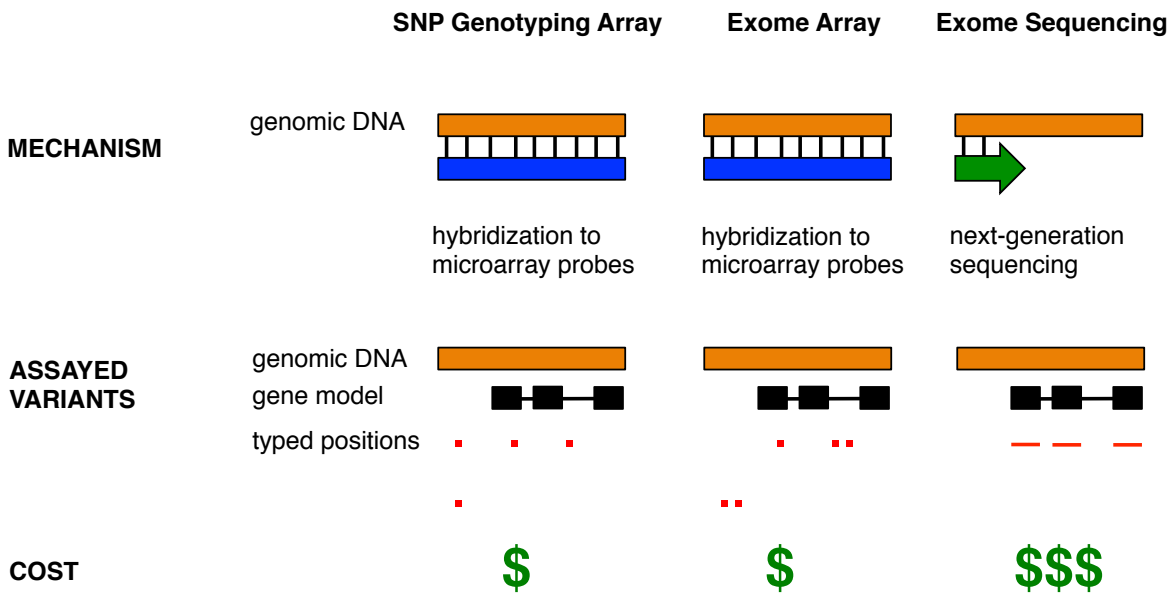


Figure 3-1: Comparison of exome array and related genotyping/sequencing technologies. Exome array serves as a bridge between conventional SNP genotyping array and exome sequencing. The exome array assays primarily variants within exonic regions of the DNA, similar to exome sequencing; however, the location of the variants must be known *a priori*. The cost of exome array is typically similar to that of other genotyping arrays, and much less than that of exome sequencing.

However, association studies are designed for linkage disequilibrium patterns among common polymorphisms, and next-generation sequencing and traditional genetic mapping approaches rely on the high penetrance of variants. Therefore, a hypothetical, moderately rare variant with moderate effect size would be too uncommon to be tagged by a standard genotyping array and have too small of an effect to be detected by linkage or genome sequencing in practical sample sizes. The exome array, or "exome chip", bridges this gap by efficiently genotyping more than 200,000 coding variants identified through sequencing studies at low cost (Figure 3-1). The exome array has been applied to phenotypes such as insulin homeostasis (Huyghe et al., 2013), bronchopulmonary dysplasia (Wang et al., 2013), and heart disease (Holmen et al., 2014; Peloso et al., 2014). For AD, Chung et al. recently reported an exome array study in Korean subjects that found association with *APOE*, *APOC1*, and *TOMM40* variants (near the

Table 3-1: Demographic Information for the Discovery Cohort

Characteristic	AD (n = 224)	Control (n = 224)	FTD (n = 168)	FTD/MND (n = 8)	PSP (n = 48)
Age, median (range), y	71 (42 to ≥89)	71 (35 to ≥89)	67 (35 to ≥89)	63 (35 to 80)	76 (55 to ≥89)
Sex, No. (%)					
Male	12 (56.7)	94 (42.0)	95 (56.5)	8 (100)	19 (39.6)
Female	97 (43.3)	130 (58.0)	73 (43.5)	0	29 (60.4)
Ancestry, No. (%)					
European	195 (87.1)	183 (81.7)	144 (85.7)	8 (100)	12 (25.0)
African American	2 (0.9)	0	0	0	0
Latino	0	4 (1.8)	1 (0.6)	0	0
Asian	20 (8.9)	27 (12.1)	9 (5.4)	0	1 (2.1)
Other	3 (1.3)	4 (1.8)	7 (4.2)	0	2 (4.2)
Unknown	4 (1.8)	6 (2.7)	7 (4.2)	0	33 (68.8)
APOE genotype, No. (%)					
E2/E2	1 (0.4)	1 (0.4)	0	0	1 (2.1)
E2/E3	7 (3.1)	19 (8.5)	16 (9.5)	1 (12.5)	3 (6.3)
E2/E4	4 (1.8)	1 (0.4)	1 (0.6)	0	2 (4.2)
E3/E3	99 (44.2)	157 (70.1)	107 (63.7)	5 (62.5)	36 (75.0)
E3/E4	92 (41.1)	40 (17.9)	40 (23.8)	1 (12.5)	6 (12.5)
E4/E4	21 (9.4)	6 (2.7)	4 (2.4)	1 (12.5)	0
Chromosome 17q21.31 haplotype, No. (%)					
H1/H1	91 (40.6)	132 (58.9)	107 (63.7)	4 (50.0)	43 (89.6)
H1/H2	48 (21.4)	52 (23.2)	33 (19.6)	3 (37.5)	5 (10.4)
H2/H2	4 (1.8)	10 (4.5)	7 (4.2)	0	0
Untyped	81 (36.2)	30 (13.4)	21 (12.5)	1 (12.5)	0

Abbreviations: AD, Alzheimer disease; FTD, frontotemporal dementia; FTD/MND, FTD with motor neuron disease; PSP, progressive supranuclear palsy.

APOE locus), but did not identify novel genetic variants (Chung et al.). Here, we report findings from the application of exome array to the multi-ethnic GIFT cohort to determine the contribution of low-frequency coding variants to susceptibility to sporadic AD, PSP, and FTD.

Subject characteristics

The initial discovery sample included 224 patients with AD, 168 patients with FTD, 8 patients with FTD with motor neuron disease (FTD/MND), 48 patients with PSP, and 224 healthy controls. Demographic characteristics are shown in Table 3-1. The ethnic makeup of this sample was predominantly Caucasian (80.7% overall). Consistent with their known roles in the respective diseases, subjects classified with AD showed high prevalence of the *APOE ε*4 allele (41.4% *ε*3/*ε*4, 9.4% *ε*4/*ε*4), and subjects classified with PSP showed high prevalence of the H1

Table 3-2: Demographic Information for the Replication Cohort

Characteristic	No. (%)			
	European (n = 135)	African American (n = 271)	Latino (n = 50)	Asian (n = 24)
Diagnosis				
AD	68 (50.4)	138 (50.9)	21 (42.0)	13 (54.2)
Control	67 (49.6)	133 (49.1)	29 (58.0)	11 (45.8)
Sex				
Male	68 (50.4)	73 (26.9)	19 (38.0)	8 (33.3)
Female	57 (42.2)	198 (73.1)	31 (62.0)	16 (66.7)
Unknown	10 (7.4)	0	0	0
Contributing center				
Emory University	21 (15.6)	223 (82.3)	0	0
University of California, Berkeley	33 (24.4)	14 (5.2)	8 (16.0)	8 (33.3)
University of California, Davis	3 (2.2)	32 (11.8)	23 (46.0)	5 (20.8)
University of California, Irvine	55 (40.7)	2 (0.7)	5 (10.0)	1 (4.2)
University of California, Los Angeles	2 (1.5)	0	0	0
University of California, San Francisco	20 (14.8)	0	0	6 (25.0)
University of Southern California	1 (0.7)	0	14 (28.0)	4 (16.7)
APOE genotype				
E2/E2	1 (0.7)	2 (0.7)	0	0
E2/E3	4 (3.0)	16 (5.9)	2 (4.0)	1 (4.2)
E2/E4	5 (3.7)	9 (3.3)	3 (6.0)	0
E3/E3	41 (30.4)	87 (32.1)	34 (68.0)	8 (33.3)
E3/E4	21 (15.6)	86 (31.7)	9 (18.0)	3 (12.5)
E4/E4	9 (6.7)	12 (4.4)	1 (2.0)	2 (8.3)
Untyped	54 (40.0)	59 (21.8)	1 (2.0)	10 (41.7)

Abbreviation: AD, Alzheimer disease.

haplotype (89.6% H1/H1, 10.4% H1/H2). The replication cohort consisted of a more ethnically heterogeneous set of patients and controls (Table 3-2).

Low-frequency exonic variants explain a fraction of the phenotypic variation in AD and FTD

The GCTA software was applied to the dataset in order to estimate the variance explained by three different classes of variants (all variants, including non-exonic variants; exonic variants only; and low-frequency exonic variants, with minor allele frequency < 5%) for each of the three diseases (AD, FTD, and PSP). In each case, a substantial portion of the observed phenotypic variance could be explained by all of the typed variants; however, due to the small sample sizes

Table 3-3: GCTA Explained Variance Analysis

	Variance Explained (SE)		
Variable	AD	FTD	PSP
All exome array variants[a]	0.44 (0.39)	0.53 (0.36)	0.57 (0.44)
Exonic fraction	0.50 (0.36)	0.45 (0.35)	0.26 (0.56)
Low-frequency exonic fraction[b]	0.41 (0.39)	0.42 (0.37)	0.03 (0.58)

Abbreviations: AD, Alzheimer disease; FTD, frontotemporal dementia; GCTA, Genome-Wide Complex Trait Analysis (http://www.complextraitgenomics.com/software/gcta/); PSP, progressive supranuclear palsy.

[a] Includes genome-wide association studies hits, HLA tag single-nucleotide polymorphisms, custom content, ancestry-informative single-nucleotide polymorphisms, and others.

[b] Less than 5% minor allele frequency between all disease cohorts and control subjects.

on which each of these estimates is based, the standard error of each measurement is high (Table 3-3).

Variant-level association testing identifies significant associations with known and novel loci

A logistic regression procedure was performed on our discovery cohort to test for association with AD, FTD, or PSP. Our method largely controlled for genomic inflation due to population stratification in each of the three disease categories (Supplementary Figure B-3), and agreed with an independent, linear mixed model method (data not shown). Two variants were suggestively associated with AD – rs769449 ($p = 1.14 \times 10^{-7}$, minor allele odds ratio = 3.0) and rs4420638 ($p = 2.58 \times 10^{-6}$, minor allele odds ratio = 2.3). Both of these variants are within the *APOE*/*TOMM40*/*APOC1* region on chromosome 19 that had been identified in previous genetic studies. One variant was associated with FTD, exm2250002 ($p = 2.08 \times 10^{-6}$, minor allele odds ratio = 0.8) corresponding to a synonymous exonic variant in the olfactory receptor genes *OR9G1* and *OR9G9*. No variants reached the suggestive p-value threshold (1×10^{-5}) in the PSP cohort. Manhattan plots depicting associations in AD, FTD, and PSP are shown in Figure 3-2.

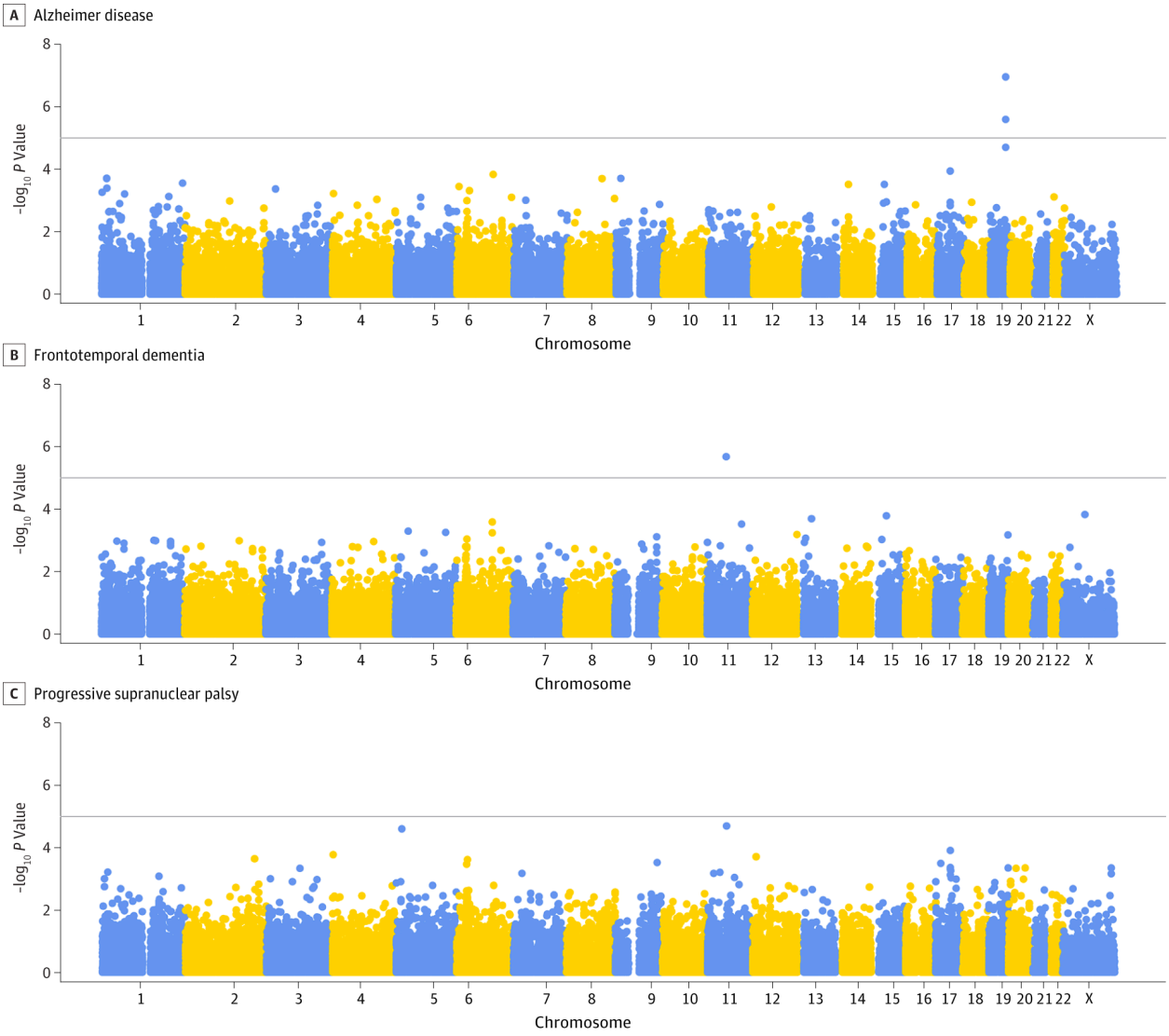


Figure 3-2: Manhattan Plot of Associations in Alzheimer Disease, Frontotemporal Dementia, and Progressive Supranuclear Palsy. The association $-\log_{10}P$ values calculated by logistic regression are presentd for for Alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy. The horizontal line indicates the suggestive *P* value threshold of $P = 1 \times 10^{-5}$. X refers to chromosome X.

Exome array genotyping replicates some previous associations found in AD, FTD, and PSP

A total of 39 polymorphisms previously associated with AD and 9 polymorphisms associated with PSP (NHGRI GWAS Catalog, http://www.genome.gov/gwastudies/, accessed April 1, 2014) were typed by the exome array. Reported susceptibility loci for FTD were not typed on this platform. We tested the association between each of these variants and their respective disease

in our cohort, as calculated by the logistic procedure described previously. For AD, the Bonferroni correction for 39 tests at a family-wise error rate of 0.05 yielded a p-value threshold at 0.0013. Two associations near *APOE*, rs2075650 ($p = 2.05 \times 10^{-5}$) and rs4420638 ($p = 2.58 \times 10^{-6}$) surpassed this pre-defined p-value threshold. While the other tested GWAS variants were not significantly associated with AD, the overall direction of association was highly consistent with previously reported results, and 23 of 32 SNPs for which the risk allele was unambiguous showed the same direction of effect as previously reported (p = 0.010, using the binomial test).

For PSP, the Bonferroni correction for 9 tests at a family-wise error rate of 0.05 yielded a p-value threshold at 0.0056. A single variant exceeded this threshold, rs8070723 (p = 0.00043) on chromosome 17, near *MAPT* (Supplementary Table B-1). Similar to the AD cohort, the direction of association was highly consistent with previously reported results, with 8 of the 9 SNPs showing the same direction of effect (p = 0.019, using the binomial test).

Gene-level testing suggests several AD candidate genes

Gene-level hypothesis testing performed using the SKAT method calculated p-values for 17,141 genes containing at least one variant that was typed by the exome array after quality control. Using a permutation procedure, an FDR of 50% was expected to be controlled at a SKAT-derived p-value of 4.54×10^{-4} for AD, 5.06×10^{-4} for FTD, and 9.65×10^{-5} for PSP. For AD, six genes exceeded this threshold, *DYSF*, *PAXIP1*, *TOP1MT*, *C3ORF1*, *SETDB1*, and *CRISPLD1*. For FTD, eight genes exceeded the threshold, *RAB21*, *AKR1B10*, *C9ORF6*, *CD5L, WDR38*, *OPHN1*, *ADORA3*, and *IKBKAP*. For PSP, two genes exceeded the threshold, *OR1Q1* and *VWA3A*. No genes were significant at an FDR threshold of 15% for any of the three diseases, AD, FTD, or PSP.

We additionally attempted to replicate the findings for AD in an additional, multi-ethnic cohort of 240 cases and 240 controls. No further samples from patients with FTD or PSP were available, so those results could not be tested. Using the Bonferroni correction, a p-value threshold of 0.0021 (considering 6 genes x 4 ethnicities for a total of 24 tests) was determined to control for a family-wise error rate of 0.05. None of the suggestive genes identified for AD were significant under this threshold in any ethnicity the replication cohort (Supplementary Table B-2). However, several genes trended toward significance in some cases, including *DYSF* in Caucasians (p = 0.076), *PAXIP1* in Latinos and East Asians (p = 0.016 and 0.037, respectively), and *TOP1MT* in African Americans (p = 0.0059). Because of previous reports of the involvement of *DYSF* and *PAXIP1* in the AD literature (see Discussion), these genes were considered interesting candidate genes for AD susceptibility. Overall, we analyzed 84 variants in *DYSF* (including 3 synonymous and 35 missense) and 5 variants in *PAXIP1* (including 1 synonymous and 4 missense) typed by the exome array, demonstrating variation in our cohort, and passing quality control criteria.

We further identified 71 candidate genes previously implicated in genetics studies of AD as categorized in the Human Gene Mutation Database version 2014.1 (Stenson et al., 2014), and extracted the gene-wise association statistics in the initial discovery set and the four replication cohorts in order to determine whether low-frequency exonic variants in these genes could contribute to AD susceptibility. Only the *ABCA7* gene (SKAT p = 0.0049) reached nominal significance. Notably, the SKAT p-value was also nominally significant in the Caucasian (p = 0.041), African American (p = 0.043), and Asian (p = 0.027) replication cohorts, but not the Latino (p = 0.61) cohort.

DYSF and PAXIP1 transcripts are differentially expressed in AD brain

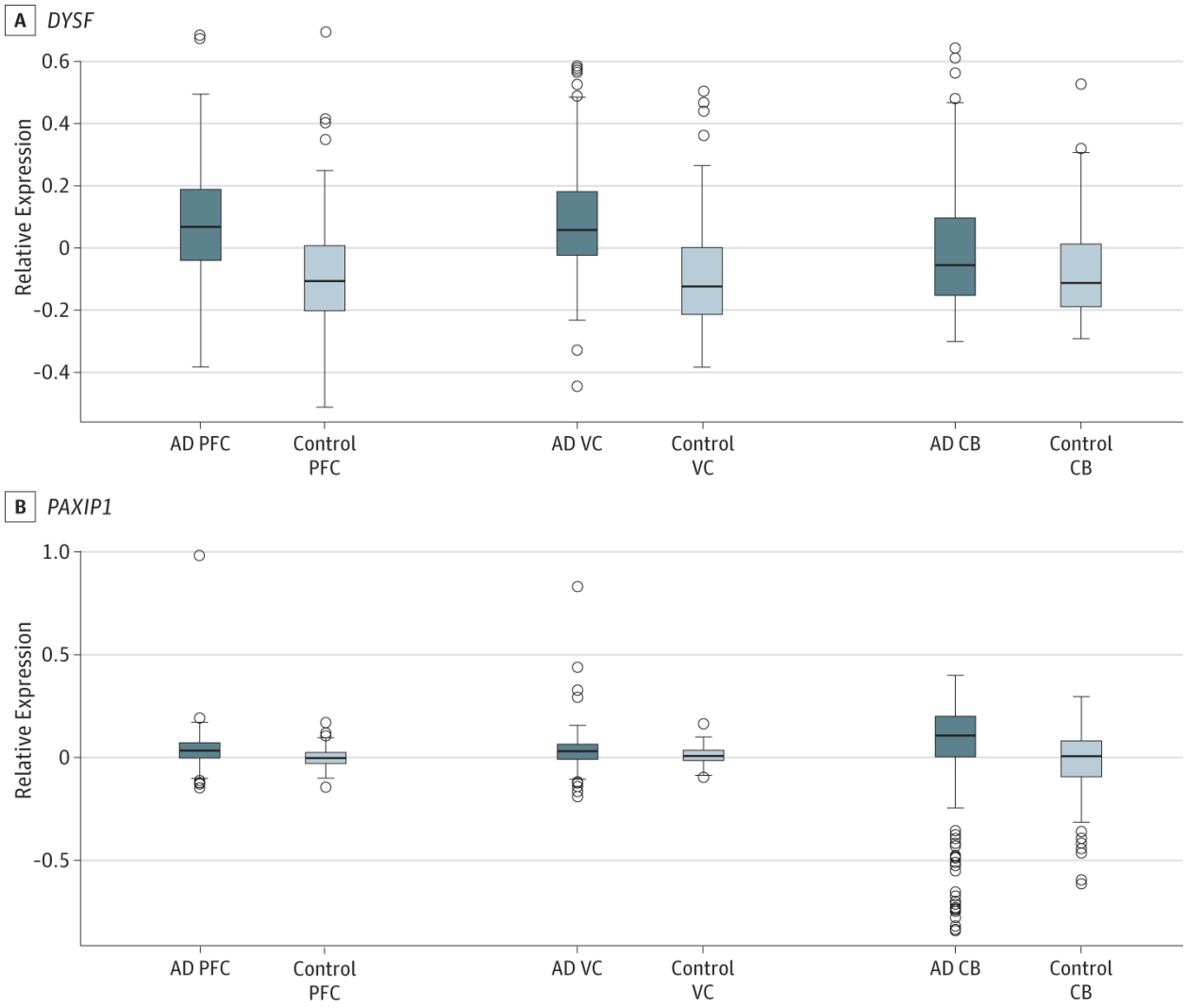


Figure 3-3: Differential Expression of *DYSF* and *PAXIP1* in Alzheimer Disease (AD) Brain. Shown is the expression of *DYSF* (A) and *PAXIP1* (B) in a public microarray data set of brain messenger RNA, grouped by brain region, in patients with AD (dark gray) vs healthy control subjects without dementia (light gray). The vertical axis represents the normalized expression residual, corrected for technical covariates. CB indicates cerebellum; PFC, prefrontal cortex; and VC, visual cortex.

To further solidify whether *DYSF* and *PAXIP1* are involved in AD pathogenesis, we examined

their relative expression levels in patients with AD and non-demented controls in a microarray dataset described in Zhang et al. (Zhang et al., 2013). The expression of *DYSF* and *PAXIP1* was significantly different between cases and controls in each of the examined brain regions (Figure 3-3). Expression of *DYSF* in the prefrontal cortex, visual cortex and cerebellum was increased in AD patients ($p < 2.2x10^{-16}$, $p = 2.33x10^{-15}$, and $p = 0.00080$, respectively). These findings were corroborated by independent data (Webster et al., 2009), which also showed increased expression of *DYSF* in AD patients' cerebral cortex ($p = 0.00023$). Similarly, expression of *PAXIP1* in the prefrontal cortex, visual cortex and cerebellum was increased in AD patients ($p = 3.6x10^{-14}$, 0.0034, and 0.00095, respectively).

Discussion

We evaluated the contribution of exonic variants to neurodegenerative disease susceptibility in a multi-ethnic cohort totaling 464 patients with AD, 168 patients with FTD, 48 patients with PSP, and 464 non-demented controls. We find that low-frequency (< 5% allele frequency) coding variants explain a sizeable proportion of the phenotypic variance in AD and FTD, although the confidence bounds for this estimate are large due to our sample size. While the well-known associations with the *APOE* locus for AD and chromosome 17q21.31 haplotype for PSP were replicated, a single novel susceptibility locus was identified: exm2250002 for FTD. Whether this variant is a true genetic signal is questionable, given that it was also the most significant signal in the PSP cohort ($p = 2.03 \times 10^{-5}$) and corresponds to a synonymous variant within *OR9G1*/*OR9G9*, members of the highly polymorphic olfactory receptor family. Gene-level testing identified suggestive signals from *DYSF* and *PAXIP1* in AD, and a trend toward significance was observed in a replication cohort in several of the tested ethnicities. A possible contribution to disease risk from exonic variants in the AD susceptibility gene *ABCA7* was also detected in multiple ethnic cohorts. We caution, however, that these results are merely suggestive and await validation in well-powered cohorts and in model systems.

The exome array's focus on coding variation, much of which has low frequency in the general population, means that large sample sizes are needed to observe statistically significant effects, unless the effect sizes are large, as is the case with the association of the *APOE ε*4 allele with AD. We estimate that a variant at 5% MAF must have over 4-fold odds ratio to achieve 80% power to identify in our AD discovery cohort. Our initial cohort of 672 patients with various forms of neurodegenerative disease and non-demented controls, and follow-up cohort of 480 patients with AD and non-demented controls, is therefore underpowered to detect associations with rare exonic variants of modest or intermediate effect sizes. Even more extreme effect sizes are required for association with FTD and PSP, which had smaller sample representation and no available replication cohort. Taken together with heritability estimates, our analyses indicate that rare variants of low or modest effect are playing a role in AD, FTD, and PSP, late onset diseases for which deleterious alleles are presumably under relatively weak selective pressure.

Furthermore, while the GIFT samples enabled testing of association in multiple ethnic groups simultaneously, our results were limited by small sample sizes in some populations. Therefore, our results do not exclude the possibility that exonic variants with lower frequency or effect size are present in the general population. In fact, the strong association with *ABCA7* (a GWAS-implicated AD susceptibility gene) by the SKAT test in several ethnic populations strongly suggests that coding variants of modest effect size within this gene are associated with AD risk. Previous GWAS have reported maximal associations with intronic polymorphisms such as rs4147929 (Lambert et al., 2013), rs115550680 (Reitz et al., 2013), rs3764650 (Hollingworth et al., 2011), and the missense polymorphism rs3752246 (Naj et al., 2011); it is possible that these variants may tag haplotypes containing causal, exonic variants. It is therefore reasonable to attempt to identify novel candidate genes containing multiple, relatively low-frequency coding variants that may contribute to AD.

While not strictly genome-wide significant, gene-wise testing results reinforce prior findings that have implicated both *DYSF* and *PAXIP1* in the pathogenesis of AD. *DYSF* encodes the protein dysferlin, and mutations in this gene are known to cause muscle diseases with an autosomal recessive pattern of inheritance, such as Miyoshi myopathy and limb-girdle muscular dystrophy type 2B, that are known as "dysferlinopathies" (Bashir et al., 1998; Liu et al., 1998). *DYSF* mRNA expression is highest in skeletal muscle, but present in other tissues such as brain, heart, and pancreas (Bashir et al., 1998; Liu et al., 1998). Dysferlin contains seven conserved C2 domains that bind Ca^{2+} and phospholipids and thereby associates with cell membranes (Abdullah et al., 2014). In skeletal muscle, dysferlin is thought play a role in Ca^{2+}-dependent sarcolemma repair by mediating fusion of repair vesicles with the injured membrane (Bansal et al., 2003; Han and Campbell, 2007) and may participate in Ca^{2+} homeostasis during mechanical stress (Kerr et al., 2014). Although its function in the central nervous system has not been extensively elaborated, dysferlin has been shown to accumulate in endothelial cells near multiple sclerosis lesions (Hochmeister et al., 2006) and within Aβ plaques of patients with AD (Galvin et al., 2006). In the latter study, Galvin et al. examined cortical regions of twelve patients (aged 85 – 99 years) at various stages of AD progression. On immunohistochemistry, dysferlin accumulated in dystrophic neurites and largely localized to Aβ plaques. The co-localization of dysferlin and Aβ-42 aggregates was also demonstrated in sporadic inclusion body myositis, suggesting that Aβ may sequester dysferlin and interfere with its normal repair functions in skeletal muscle (Cacciottolo et al., 2013).

The second highlighted gene, *PAXIP1*, encodes for a nuclear protein with six BRCT domains, hinting at its function in DNA repair pathways (Jowsey et al., 2004). PAXIP1 may participate in ATM-mediated activation of p53, demonstrated by binding with 53BP1 at sites of DNA damage (Jowsey et al., 2004; Munoz et al., 2007; Wu et al., 2009), sensitization to ionizing radiation-

induced damage following *in vitro* knockdown (Jowsey et al., 2004), and involvement in cell cycle progression (Cho et al., 2003) and lymphocyte development (Daniel et al., 2010). Though variants in *PAXIP1* have not been definitively associated with disease, Rademakers et al. identified a significant linkage peak at 7q36 in a large, three-generation Dutch pedigree with multiplex AD (Rademakers et al., 2005). The risk allele of the *D7S798* marker also appeared to increase AD risk by 2.7 times in a Dutch population-based cohort. Follow-up sequencing of the coding exons of 29 candidate genes revealed only a single rare variant, a synonymous Ala626 mutation in *PAXIP1*.

Interestingly, to our knowledge the neuropathological findings of Galvin et al. (2006) and the linkage study of Rademakers et al. (2005) are currently the only publications that implicate *DYSF* and *PAXIP1* in AD pathogenesis. Our analysis of published microarray studies indicated widespread increases in *DYSF* and *PAXIP1* mRNA expression in brain regions of AD patients. However, these results do not provide direct evidence of the genes' roles in AD. The results presented here by exome array genotyping add additional support for the causal pathogenicity of *DYSF* and *PAXIP1* in the general population. Although we could not ascertain whether any of the assayed variants directly affected expression of *DYSF* and *PAXIP1*, the fact that these genes were both identified by exome array analysis and by differential expression analysis provides convergent streams of evidence for their involvement in AD. Besides partial, nominal replication within our cohort, our findings are further corroborated by a recently published exome chip study in AD. Chung et al. (2014) found a strong (but not genome-wide significant) association for *DYSF* ($p = 1.6x10^{-5}$) with AD using the SKAT-O test in a Korean cohort; the association with *PAXIP1* was not reported. The overlap with our suggestive results indicates a high prior probability for the pathogenicity of variants in *DYSF* (and possibly also *PAXIP1*), and follow-up studies are warranted.

The overall genetic architecture of neurodegenerative diseases is complex and just beginning to be clearly defined. Our work has strengthened the case for two AD candidate genes and provides one of the first glimpses at this genetic variation that heretofore had not been widely studied. We hope that the results described herein will provide useful insight into the genetics of AD, FTD, and PSP; that the data will provide a valuable multi-ethnic cohort with exome array genotyping data for future studies and meta-analyses in each of the three diseases; and in the long term, that increased understanding of the genetic underpinnings will lead to improvements in diagnosis and management for patients suffering from neurodegenerative diseases.

Methods

Study cohort. Patients and healthy controls were enrolled as part of the Genetic Investigation in FrontoTemporal Dementia (GIFT), a study of the genetics of neurodegenerative disease at the UCSF Memory and Aging Center (Coppola et al., 2007; Li et al., 2014). An additional 32 DNA samples from patients with PSP were extracted from post-mortem brain tissue from the New York Brain Bank. A subset of these subjects were initially selected for genotyping for this study (Table 3-1). Patients diagnosed with FTD/MND were excluded from further analysis due to small sample size and potential genetic heterogeneity.

Replication cohort. As part of the GIFT study, subjects were also enrolled from other sites, including Emory University, the University of Southern California, UC-Berkeley, UC-Davis, UC-Irvine, and UCLA. This cohort was multi-ethnic, including individuals of self-identified Caucasian, African-American, Latino, and East Asian ancestry. Following initial data analysis, 480 subjects from this additional group of patients, including 240 patients with AD and 240 non-demented controls, were genotyped on the Illumina Infinium HumanExome BeadChip Kit as previously described (Table 3-2). These subjects were analyzed as above, but due to genetic heterogeneity, were divided into four general groups (Caucasian, African-American, Latino, and

East Asian) based on self-reported ancestry. To ensure proper classification and minimize the inclusion of misplated samples, ancestry was estimated by multidimensional scaling in PLINK, using the entire set of genotyped variants by exome array (Supplementary Figure B-2). Following this procedure, 44 samples were suspected of misclassification and were removed from further analysis.

Exome array genotyping. Exonic and non-exonic variants were genotyped by the Illumina Infinium HumanExome BeadChip Kit. While mostly consisting of coding variants from prior sequencing studies, the exome arrays also included markers for previously described GWAS hits, ancestry-informative markers, randomly selected synonymous variants, HLA tag SNPs, and others (Huyghe et al., 2013), in total comprising 250,272 genotyped markers per sample. All data was processed using GRCh37/hg19 coordinates. Quality control procedures were enacted to remove suspect variants and minimize the effect of population structure on data analysis. See eAppendix and eFigure 1 for further details on genotyping and data pre-processing procedures.

Statistical analysis. The total phenotypic (disease) variance explained by the genotyped variants was determined using a restricted maximum likelihood model implemented in GCTA. Variant-level association with AD, FTD, and PSP was tested using a logistic regression model that corrected for population structure. Association on the gene level was tested using the Sequence Kernel Association Test (Wu et al., 2011), a non-burden test that is sensitive in the presence of neutral genetic variants. Genes that showed suggestive associations with AD were also tested in previously described brain mRNA expression datasets (Webster et al., 2009; Zhang et al., 2013). See Appendix Bfor a more detailed description of the statistical methods used.

References

Abdullah, N., Padmanarayana, M., Marty, Naomi J., and Johnson, Colin P. (2014). Quantitation of the Calcium and Membrane Binding Properties of the C2 Domains of Dysferlin. Biophysical Journal *106*, 382-389.

Baker, M., Litvan, I., Houlden, H., Adamson, J., Dickson, D., Perez-Tur, J., Hardy, J., Lynch, T., Bigio, E., and Hutton, M. (1999). Association of an Extended Haplotype in the Tau Gene with Progressive Supranuclear Palsy. Human Molecular Genetics *8*, 711-715.

Bansal, D., Miyake, K., Vogel, S.S., Groh, S., Chen, C.-C., Williamson, R., McNeil, P.L., and Campbell, K.P. (2003). Defective membrane repair in dysferlin-deficient muscular dystrophy. Nature *423*, 168-172.

Bashir, R., Britton, S., Strachan, T., Keers, S., Vafiadaki, E., Lako, M., Richard, I., Marchand, S., Bourg, N., Argov, Z*., et al.* (1998). A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. Nature Genetics *20*, 37-42.

Cacciottolo, M., Nogalska, A., D'Agostino, C., Engel, W.K., and Askanas, V. (2013). Dysferlin is a newly identified binding partner of AβPP and it co-aggregates with amyloid-β42 within sporadic inclusion-body myositis (s-IBM) muscle fibers. Acta Neuropathologica *126*, 781-783.

Cho, E.A., Prindle, M.J., and Dressler, G.R. (2003). BRCT Domain-Containing Protein PTIP Is Essential for Progression through Mitosis. Molecular and Cellular Biology *23*, 1666-1673.

Chung, S.J., Kim, M.-J., Kim, J., Kim, Y.J., You, S., Koh, J., Kim, S.Y., and Lee, J.-H. (2014). Exome array study did not identify novel variants in Alzheimer's disease. Neurobiology of Aging *35*, 1958.e1913-1958.e1914.

Coppola, G., Chinnathambi, S., Lee, J.J., Dombroski, B.A., Baker, M.C., Soto-Ortolaza, A.I., Lee, S.E., Klein, E., Huang, A.Y., Sears, R*., et al.* (2012). Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. Human Molecular Genetics *21*, 3500-3512.

Coppola, G., Miller, B.L., Chui, H., Varpetian, A., Levey, A., Cotman, C.W., DeCarli, C., Mendez, M.F., Bartzokis, G., and Kukull, W.A. (2007). Genetic investigation in frontotemporal dementia and Alzheimer's disease: the GIFT Study. Annals of Neurology *62*, S52.

Daniel, J.A., Santos, M.A., Wang, Z., Zang, C., Schwab, K.R., Jankovic, M., Filsuf, D., Chen, H.-T., Gazumyan, A., Yamane, A*., et al.* (2010). PTIP Promotes Chromatin Changes Critical for Immunoglobulin Class Switch Recombination. Science *329*, 917-923.

Ferrari, R., Hernandez, D.G., Nalls, M.A., Rohrer, J.D., Ramasamy, A., Kwok, J.B.J., Dobson-Stone, C., Brooks, W.S., Schofield, P.R., Halliday, G.M.*, et al.* (2014). Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neurology *13*, 686-699.

Galvin, J., Palamand, D., Strider, J., Milone, M., and Pestronk, A. (2006). The muscle protein dysferlin accumulates in the Alzheimer brain. Acta Neuropathologica *112*, 665-671.

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S.K., Younkin, S.*, et al.* (2013). TREM2 Variants in Alzheimer's Disease. New England Journal of Medicine *368*, 117-127.

Han, R., and Campbell, K.P. (2007). Dysferlin and muscle membrane repair. Current Opinion in Cell Biology *19*, 409-416.

Hochmeister, S., Grundtner, R., Bauer, J., Engelhardt, B., Lyck, R., Gordon, G., Korosec, T., Kutzelnigg, A., Berger, J.J., Bradl, M.*, et al.* (2006). Dysferlin Is a New Marker for Leaky Brain Blood Vessels in Multiple Sclerosis. Journal of Neuropathology and Experimental Neurology *65*, 855-865.

Hoglinger, G.U., Melhem, N.M., Dickson, D.W., Sleiman, P.M.A., Wang, L.-S., Klei, L., Rademakers, R., de Silva, R., Litvan, I., Riley, D.E.*, et al.* (2011). Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nature Genetics *43*, 699-705.

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.C., Carrasquillo, M.M., Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskvina, V.*, et al.* (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nature Genetics *43*, 429.

Holmen, O.L., Zhang, H., Zhou, W., Schmidt, E., Hovelson, D.H., Langhammer, A., Løchen, M.-L., Ganesh, S.K., Mathiesen, E.B., Vatten, L.*, et al.* (2014). No large-effect low frequency coding variation found for myocardial infarction. Human Molecular Genetics.

Huyghe, J.R., Jackson, A.U., Fogarty, M.P., Buchkovich, M.L., Stancakova, A., Stringham, H.M., Sim, X., Yang, L., Fuchsberger, C., Cederberg, H.*, et al.* (2013). Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nature Genetics *45*, 197-201.

Jowsey, P.A., Doherty, A.J., and Rouse, J. (2004). Human PTIP Facilitates ATM-mediated Activation of p53 and Promotes Cellular Resistance to Ionizing Radiation. Journal of Biological Chemistry *279*, 55562-55569.

Kerr, J.P., Ward, C.W., and Bloch, R.J. (2014). Dysferlin at transverse tubules regulates Ca2+ homeostasis in skeletal muscle. Front Physiol *5*, 1-5.

Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W.*, et al.* (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature Genetics *45*, 1452-1458.

Lendon, C.L., Ashall, F., and Goate, A.M. (1997). Exploring the etiology of alzheimer disease using molecular genetics. JAMA *277*, 825-831.

Li, Y., Chen, J.A., Sears, R.L., Gao, F., Klein, E.D., Karydas, A., Geschwind, M.D., Rosen, H.J., Boxer, A.L., Guo, W.*, et al.* (2014). An Epigenetic Signature in Peripheral Blood Associated with the Haplotype on 17q21.31, a Risk Factor for Neurodegenerative Tauopathy. PLoS Genetics *10*, e1004211.

Liu, J., Aoki, M., Illa, I., Wu, C., Fardeau, M., Angelini, C., Serrano, C., Urtizberea, J.A., Hentati, F., Hamida, M.B.*, et al.* (1998). Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. Nature Genetics *20*, 31-36.

Munoz, I.M., Jowsey, P.A., Toth, R., and Rouse, J. (2007). Phospho-epitope binding by the BRCT domains of hPTIP controls multiple aspects of the cellular response to DNA damage. Nucleic Acids Research *35*, 5312-5322.

Naj, A.C., Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buros, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K.*, et al.* (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nature Genetics *43*, 436-441.

Peloso, G.M., Auer, P.L., Bis, Joshua C., Voorman, A., Morrison, Alanna C., Stitziel, Nathan O., Brody, Jennifer A., Khetarpal, Sumeet A., Crosby, Jacy R., Fornage, M.*, et al.* (2014). Association of Low-Frequency and Rare Coding-Sequence Variants with Blood Lipids and Coronary Heart Disease in 56,000 Whites and Blacks. American Journal of Human Genetics *94*, 223-232.

Pericak-Vance, M., Bebout, J., Gaskell, P., Yamaoka, L., Hung, W.-Y., Alberts, M., Walker, A., Bartlett, R., Haynes, C., and Welsh, K. (1991). Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. American Journal of Human Genetics *48*, 1034.

Pittman, A.M., Myers, A.J., Abou-Sleiman, P., Fung, H.C., Kaleem, M., Marlowe, L., Duckworth, J., Leung, D., Williams, D., Kilford, L.*, et al.* (2005). Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. Journal of Medical Genetics *42*, 837-846.

Rademakers, R., Cruts, M., Sleegers, K., Dermaut, B., Theuns, J., Aulchenko, Y., Weckx, S., De Pooter, T., Van den Broeck, M., Corsmit, E.*, et al.* (2005). Linkage and Association Studies Identify a Novel Locus for Alzheimer Disease at 7q36 in a Dutch Population-Based Sample. American Journal of Human Genetics *77*, 643-652.

Rademakers, R., Neumann, M., and Mackenzie, I.R. (2012). Advances in understanding the molecular basis of frontotemporal dementia. Nat Rev Neurol *8*, 423-434.

Reitz, C., Jun, G., Naj, A., and et al. (2013). Variants in the ATB-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ4, and the Risk of Late-Onset Alzheimer Disease in African Americans. JAMA *309*, 1483-1492.

Stenson, P., Mort, M., Ball, E., Shaw, K., Phillips, A., and Cooper, D. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Human Genetics *133*, 1-9.

Van Deerlin, V.M., Sleiman, P.M.A., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.-S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M.*, et al.* (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nature Genetics *42*, 234-239.

Wang, H., St. Julien, K.R., Stevenson, D.K., Hoffmann, T.J., Witte, J.S., Lazzeroni, L.C., Krasnow, M.A., Quaintance, C.C., Oehlert, J.W., Jelliffe-Pawlowski, L.L.*, et al.* (2013). A Genome-Wide Association Study (GWAS) for Bronchopulmonary Dysplasia. Pediatrics *132*, 290-297.

Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M.*, et al.* (2009). Genetic Control of Human Brain Transcript Expression in Alzheimer Disease. American Journal of Human Genetics *84*, 445-458.

Wu, J., Prindle, M.J., Dressler, G.R., and Yu, X. (2009). PTIP Regulates 53BP1 and SMC1 at the DNA Damage Sites. Journal of Biological Chemistry *284*, 18078-18084.

Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. American Journal of Human Genetics *89*, 82-93.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, Alexei A., Zhang, C., Xie, T., Tran, L., Dobrin, R.*, et al.* (2013). Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. Cell *153*, 707-720.

Chapter 4: Genome Sequencing and Rare Variation in Progressive Supranuclear Palsy

Introduction

Progressive supranuclear palsy (PSP) is a disease characterized by tau aggregates in neuropathology. PSP patients classically exhibit the eponymous supranuclear gaze palsy in the vertical plane, and suffer from parkinsonian symptoms such as frequent falls (Williams and Lees, 2009). In this regard, PSP is a prototypical tauopathy. Its clinical presentation is highly correlated with its characteristic neuropathology, prominent tau aggregation with absent involvement of other pathogenic proteins (e.g. amyloid plaques of Alzheimer's disease). A unique risk factor also implies high genetic homogeneity: nearly all PSP patients possess two copies of the H1 allele at the chromosome 17q21.31 inversion haplotype (Pittman et al., 2004). For this reason, it is the ideal disease to study the genetics of tauopathies.

We previously reported GWAS and exome array studies in PSP (Chen et al., 2015). Whole-genome sequencing (WGS) is currently the most detailed glimpse at rare and common variation available. Compared to exome sequencing, WGS provides more uniform and complete coverage over the exome, sequencing of non-coding regions, and the ability to resolve structural variation. Here, we describe preliminary findings in genome sequencing of PSP and outline a path for future characterization of the genetics of the disease.

Subject characteristics

Subjects were sequenced in two batches by Illumina ("Illumina cohort") and the New York Genome Center ("NYGC cohort") (Figure 4-1). Most of the downstream analyses were performed on the Illumina cohort only; because of slight differences in the sequencing batch, cross-platform results were not compatible for joint analysis. The Illumina cohort consisted of 276 patients with PSP and 88 healthy elderly controls.

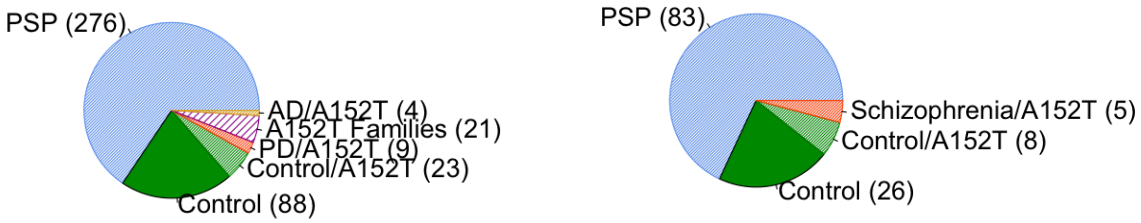


Figure 4-1: Sequencing cohorts by batch. Left: Illumina cohort, sequenced on the Illumina HiSeq 2500 and with downstream analysis from a proprietary Illumina pipeline. Right: NYGC cohort, sequenced on the Illumina HiSeqX and with downstream analysis using the standard GATK pipeline.

MAPT A152T: A rare, exonic variant associated with multiple neurodegenerative diseases

The *MAPT* A152T mutation was one of the initial rare exonic mutations identified by our group as a risk factor in multiple neurodegenerative diseases, including PSP, AD, and FTD (Coppola et al., 2012). In the Illumina cohort, we identified a startling 2% allele frequency of *MAPT* A152T among PSP patients. We confirmed this result in a larger cohort of 3100 patients with neurodegenerative disease and 4351 controls recruited worldwide across collaborating centres, including: (i) University of California San Francisco and the multicentre Davunetide trial (Boxer et al., 2014) (Allon series); (ii) Gladstone Institute (Gladstone Turkish series); (iii) University of Brescia and San Raffaele Scientific Institute (Italian series); (iv) Northwestern University (Northwestern series); (v) Rush Alzheimer's Disease Center (Religious Orders Study and Rush Memory and Aging Project); (vi) University of California Los Angeles (Small series); (vii) Instituto de Salud Carlos III, Madrid, Spain (Spanish series); (viii) University of Toronto Memory Clinic (Toronto series); (ix) School of Medicine, Yale University (Turkish series); (x) University of California San Francisco Memory and Aging Center (UCSF series); and (xi) other centres (Other series). All individuals gave authorization for genetic testing research in accordance with the local regulations. Each institution's Committee on Human Research approved the study. Clinical diagnoses were rendered by expert neurologists at each institution. In this study, a diagnosis of FTD denotes clinically defined behavioural variant FTD or primary progressive aphasia. Participants or their surrogates provided informed consent before participation.

Cohort	#Carriers	#Individuals	%Carriers	Odds Ratio	*P*−value
ALS	0	24	0.00	0.00	−
PD	0	206	0.00	0.00	−
CBS	1	133	0.75	3.29	0.2822
MCI	2	462	0.43	1.89	0.3225
AD	5	927	0.54	2.35	0.1621
FTD	7	913	0.77	3.35	0.0180
PSP−S	8	435	1.84	8.13	8.43E−05
Control	10	4351	0.23		
Overall	**23**	**3100**	**0.74**	**3.24**	**0.0013**

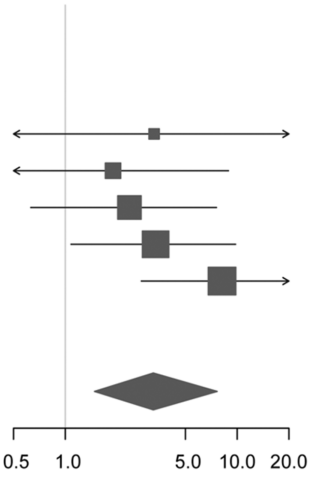


Figure 4-2: MAPT p.A152T carrier frequencies and associated odds ratio for the different disease cohorts. The total number of individuals and p.A152T carriers for each of the disease cohorts and controls is shown in the table (left), with odds ratios and nominal P-values depicted in the forest plot (right). Overall refers to the combined neurological disease patient samples. In the forest plot, squares represent the estimate odds ratio and are drawn proportional to the weight of the sample and lines represent 95% confidence intervals.

Among 3100 patients with neurodegenerative disease, we identified 23 p.A152T carriers (Figure 4-2): one carrier in 133 cases with corticobasal syndrome (0.75%); two carriers in 462 cases with mild cognitive impairment (0.43%); five carriers in 927 cases with Alzheimer's disease (0.54%); seven carriers in 913 FTD cases (0.77%); and eight in 435 cases with PSP-S (1.84%). No carrier was identified among our cohorts of amyotrophic lateral sclerosis (n = 24) and Parkinson's disease (n = 206) individuals. Ten in 4351 control individuals carried the p.A152T allele (0.23%). This frequency is similar to the one observed in over 60 000 unrelated individuals from the Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org/), with 159 carriers (three of them homozygous) among 60 472 individuals (0.26%).

A combined analysis performed on the 3100 patients placed the estimated OR at 3.24 (CI: 1.48–7.65, Fisher's exact test, P = 0.0013) for overall neurodegenerative diseases versus controls. Analyses on the different individual disease cohorts, revealed OR ranging between 1.89 (in the MCI cohort) and 8.13 (in the PSP-S cohort) (Figure 4-2). However, the association

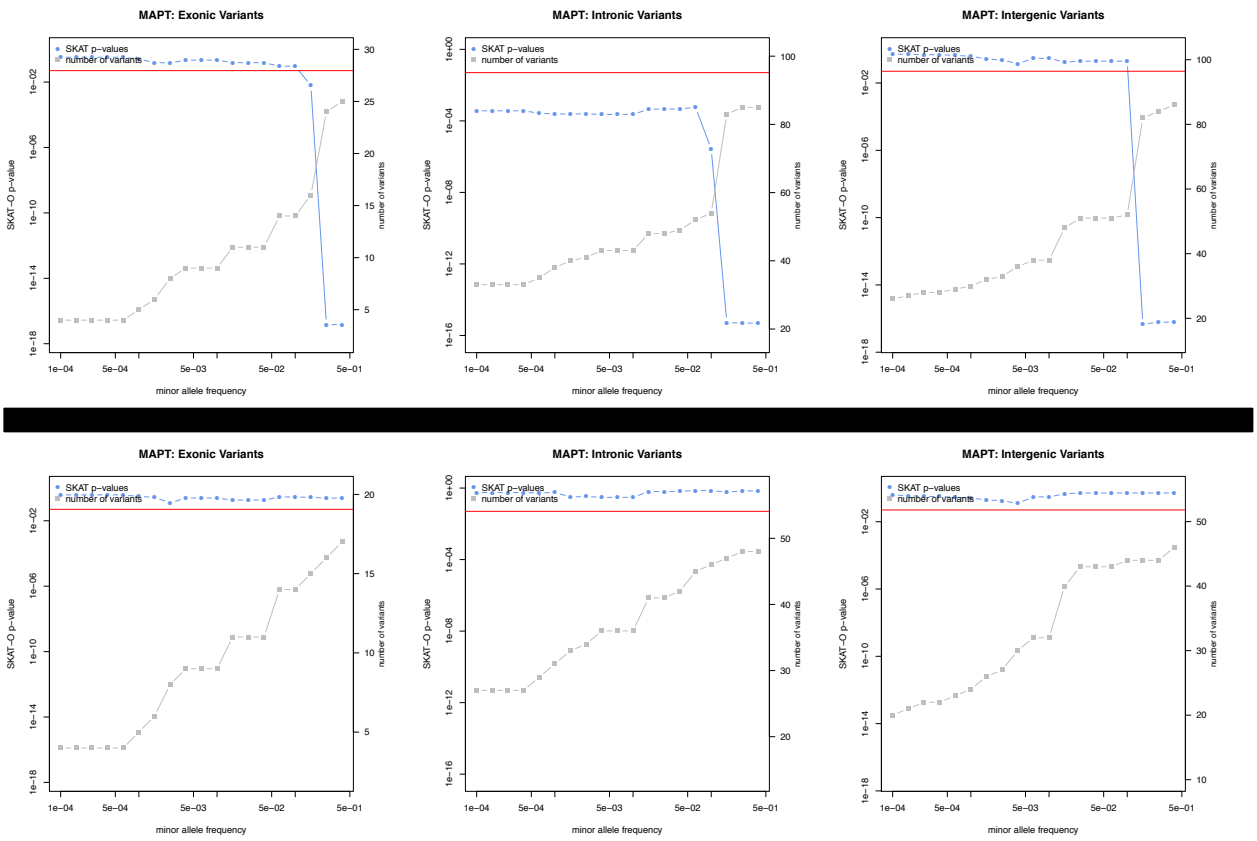


Figure 4-3: Sequence kernel association test (SKAT) p-value for variant classes within *MAPT*. The SKAT p-value (blue line, left vertical axis) as a function of maximal minor allele frequency and the total number of variants falling in each class (grey line, right vertical axis) for each class of variants, exonic, intronic, and intergenic (promoter). A nominally significant p-value (p = 0.05) is demarcated by the red line. This analysis was performed for all samples (top) and H1/H1 samples only (bottom).

was only significant for the PSP-S (OR = 8.13, CI: 2.77–22.99, Fisher's exact test, P = 8.43 × 10^{-5}) cohort and nominally significant for the FTD (OR = 3.35, CI: 1.08–9.79, Fisher's exact test, P = 0.0180) cohort. As this is a targeted replication study aiming to test a single variant, these P-values are sufficient support for the hypothesis, as they do not require correction for testing of multiple loci.

Gene-wise Variant Analysis in PSP GWAS regions

Because we lacked statistical power to detect variant burden in genes genome-wide, we focused the search space to genes identified in a prior genome-wide association study of PSP (Hoglinger et al., 2011). The genetic architecture of other neurodegenerative diseases, including

Parkinson's disease, includes causal rare variants that overlap GWAS signals (Singleton et al., 2013). In PSP, rare variants in *MAPT* also overlap two independent associations with common variants (the chr17q21.31 haplotype and rs242557) (Hoglinger et al., 2011). We therefore reasoned that an abnormal distribution of variants might be detectable from genes implicated in PSP GWAS signals.

Looking first at the *MAPT* gene, we subdivided variants into three categories: exonic, intronic, and near intergenic (i.e., promoter). At various levels of maximum minor allele frequency (e.g. ranging from rare variants only to all variants), we used the sequence kernel association test (SKAT) (Lee et al., 2012) to identify differences in distribution of variant classes between the PSP patients and controls. For both classes, a significantly increased burden of variants in tau was observed in PSP patients (Figure 4-3). However, this effect disappeared when restricting to H1/H1 subjects only, suggesting that the entire enrichment was due to the effect of the haplotype. The other GWAS genes showed similar distributions of variants in both cases and controls; we did not identify a statistically significant signal from any of the tested genes, even at nominal significance.

Recurrent loss-of-function variants are enriched in PSP patients

Turning our attention genome-wide, we attempted to stratify variants with the most deleterious functional effects. These variants should be highly enriched for downstream genomic consequences and provide the best opportunity to identify genetic enrichments. We further included two additional filters to improve the enrichment further: the haploinsufficiency score for each gene (Huang et al., 2010), a proxy for the likelihood that loss of the gene would be deleterious; and the number of recurrences of rare variants in the sample cohort. Strikingly, significant enrichment was observed for these highly selected variant classes, particularly those genes that are predicted to be highly deleterious by the haploinsufficiency score (Figure 4-4).

PSP recur.	Total	HI > 0.1	HI > 0.2	HI > 0.3	HI > 0.4	HI > 0.5	HI > 0.6	HI > 0.7	HI > 0.8	HI > 0.9
≥1	1648 (p=0.027)	1170 (p=0.047)	653 (p=0.055)	428 (p=0.162)	307 (p=0.090)	213 (p=0.110)	161 (p=0.023)	124 (p=0.004)	79 (p=0.055)	44 (p=0.090)
≥2	323 (p=0.002)	230 (p=0.007)	131 (p=0.015)	86 (p=0.045)	60 (p=0.106)	38 (p=0.205)	28 (p=0.206)	20 (p=0.069)	15 (p=0.091)	10 (p=0.057)
≥3	117 (p=3e-04)	89 (p=0.003)	46 (p=0.019)	22 (p=0.450)	15 (p=0.492)	11 (p=0.337)	7 (p=0.466)	6 (p=0.287)	5 (p=0.260)	4 (p=0.093)
≥4	52 (p = 0.001)	43 (p=0.003)	22 (p=0.055)	9 (p=0.603)	8 (p = 0.365)	5 (p = 0.466)	4 (p=0.444)	3 (p=0.476)	3 (p = 0.275)	3 (p=0.176)
≥5	22 (p=0.035)	17 (p=0.086)	9 (p=0.304)	4 (p=0.739)	4 (p=0.527)	3 (p=0.424)	2 (p=0.532)	1 (p=0.698)	1 (p=0.628)	1 (p=0.483)

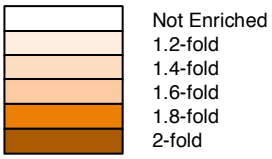


Figure 4-4: Enrichment of loss-of-function variants in PSP patients. Loss-of-function variants were stratified by haploinsufficiency score (columns) and the number of recurrences within the dataset (horizontal axis).

The genes in significant enriched categories were themselves enriched in neuronal-related Gene Ontology categories, further suggesting that these may represent disease-causal pathogenic variants.

Discussion

Whole genome sequencing is an incredibly rich dataset for identification of new genetic risk factors, but the interpretation is complicated by uncertainty in variant annotation and poor statistical power. Although we have not identified new risk genes in PSP, we find many suggestions that such a signal exists in the data. First, looking within the tau gene, we found a massive overrepresentation of the *MAPT* A152T allele that was previously identified by our group (Coppola et al., 2012). Expanding this confirmation study to a large, multinational cohort spanning multiple diseases, we attempted to clarify the situation with regard to FTD spectrum and other clinical syndromes known to be caused by FTLD-tau. In the present study, we find significant risk associations in our independent FTD and PSP-S cohorts with the p.A152T variant. The groups of Angeleen Fleming and David Rubinsztein further worked on the mechanism of the mutation, raising intriguing possibilities into tau biology (Lopez et al., 2017).

Enrichment of multiple-hit deleterious variants in highly probable haploinsufficient genes in PSP patients yields yet more evidence that a signal exists. However, we did not identify individual genes harboring loss-of-function variants, even when focusing on the GWAS regions to increase statistical power. Other approaches may be required to identify potential candidate genes of interest; for example, the genes overlapping PSP loci may not be the ultimate endpoint of the association signal.

Future work will extend the findings in the study by rigorously combining the sequenced sample cohorts to permit direct joint analysis. Furthermore, additional types of genetic variation, such as copy number variants, will be ascertained. New analysis methods may decipher the potential regulatory functions of non-coding variation and detect association with PSP risk, particularly focusing in the tau region. Finally, improvements in statistical power from increased sample cohorts may also improve the analysis.

Methods

Genome sequencing. Blood samples were obtained from the Allon davunetide trial (Boxer et al., 2014), which featured deep clinical phenotyping of subjects diagnosed with PSP. Controls, and additional cases, were obtained from patients seen at the University of California, San Francisco Memory and Aging Center. Genome sequencing was performed in multiple batches at Illumina and the New York Genome Center.

Variant discovery from whole genome sequencing. To minimize the effect of sequencing batch on variant calls, genomes were reprocessed from raw files. Variant re-calling was performed using a standard pipeline. Briefly, BAM files from each center (Illumina pipelines and New York Genome Center) were converted to the original FASTQ format. Realignment to the reference genome was performed with SNAP and ADAM (Zaharia et al., 2011). Single nucleotide variants and short indels were called using the Genome Analysis Toolkit (GATK). Structural variants

were called using GenomeSTRiP; however, the results were preliminary and not included in this analysis. Variant annotation was performed using ANNOVAR (Wang et al., 2010).

Confirmation of the MAPT A152T allele. Genotypes were obtained using TaqMan® SNP assays from Life Technologies on a LightCycler® 480 System. A custom assay (#AHHR7R6) was designed for *MAPT* p.A152T rs143624519. Forward primer sequence was CCAATGGTGAAAAACCCCTCTATCA and reverse primer sequence was TTGGCCTGGCCCTTCTG. Reporter sequences were AAAACGAAGATCACCACACC and ACGAAGATCGCCACACC. p.A152T carriers were confirmed using Sanger sequencing. Statistical analysis was performed in R (version 3.1.3, www.r-project.org).

References

Boxer, A.L., Lang, A.E., Grossman, M., Knopman, D.S., Miller, B.L., Schneider, L.S., Doody, R.S., Lees, A., Golbe, L.I., Williams, D.R*., et al.* (2014). Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. The Lancet Neurology *13*, 676-685.

Chen, J.A., Wang, Q., Davis-Turak, J., and et al. (2015). A multiancestral genome-wide exome array study of alzheimer disease, frontotemporal dementia, and progressive supranuclear palsy. JAMA Neurology *72*, 414-422.

Coppola, G., Chinnathambi, S., Lee, J.J., Dombroski, B.A., Baker, M.C., Soto-Ortolaza, A.I., Lee, S.E., Klein, E., Huang, A.Y., Sears, R*., et al.* (2012). Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. Human Molecular Genetics *21*, 3500-3512.

Hoglinger, G.U., Melhem, N.M., Dickson, D.W., Sleiman, P.M.A., Wang, L.-S., Klei, L., Rademakers, R., de Silva, R., Litvan, I., Riley, D.E*., et al.* (2011). Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nature Genetics *43*, 699-705.

Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and Predicting Haploinsufficiency in the Human Genome. PLoS Genetics *6*, e1001154.

Lee, S., Wu, M.C., and Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. Biostatistics *13*, 762-775.

Lopez, A., Lee, S.E., Wojta, K., Ramos, E.M., Klein, E., Chen, J., Boxer, A.L., Gorno-Tempini, M.L., Geschwind, D.H., Schlotawa, L*., et al.* (2017). A152T tau allele causes neurodegeneration that can be ameliorated in a zebrafish model by autophagy induction. Brain *140*, 1128-1146.

Pittman, A.M., Myers, A.J., Duckworth, J., Bryden, L., Hanson, M., Abou-Sleiman, P., Wood, N.W., Hardy, J., Lees, A., and de Silva, R. (2004). The structure of the tau haplotype in controls and in progressive supranuclear palsy. Human Molecular Genetics *13*, 1267-1274.

Singleton, A.B., Farrer, M.J., and Bonifati, V. (2013). The genetics of Parkinson's disease: Progress and therapeutic implications. Movement Disorders *28*, 14-23.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research *38*, e164.

Williams, D.R., and Lees, A.J. (2009). Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges. The Lancet Neurology *8*, 270-279.

Zaharia, M., Bolosky, W.J., Curtis, K., Fox, A., Patterson, D., Shenker, S., Stoica, I., Karp, R.M., and Sittler, T. (2011). Faster and more accurate sequence alignment with SNAP. arXiv preprint arXiv:11115572.

Chapter 5: Methylation in Alzheimer's disease, frontotemporal dementia, and progressive supranuclear palsy

Introduction

Epigenetics is one of the most rapidly expanding fields in biology, and is uncovering additional levels of complexity in the human genome, including DNA methylation, histone modifications, and intra- and inter-chromosomal interactions mediated by chromatin proteins (Feinberg, 2010; Portela and Esteller, 2010). Changes in methylation represent a key area where environmental factors can modify or interact with inherited genetic factors (DNA sequence) to alter the functional output of the genome. Disease-causing genes involved in epigenetic modifications have been identified, most notably for neurodevelopmental disorders such as Rett syndrome (Zoghbi, 2009). A very limited number of studies have addressed specific epigenetic modifications relevant to neurological diseases and dementia (Akbarian et al., 2013; Jakovcevski and Akbarian, 2012; Lu et al., 2013; Urdinguio et al., 2009). Additionally, epigenetic signatures have been reported for different brain regions (Ladd-Acosta et al., 2007; van Eijk et al., 2012), for regional brain aging (Hernandez et al., 2011), and aging in general (Horvath, 2013) further supporting epigenetic studies in patients with neurodegenerative diseases.

Progressive supranuclear palsy (PSP) is a neurodegenerative disease typically characterized by parkinsonism, postural instability, and cognitive impairment (Steele et al., 1964). Pathologically, PSP is defined by the accumulation of tau protein in subcortical and cortical regions (Williams and Lees, 2009), showing substantial overlap with other neurodegenerative diseases characterized by tau accumulation and grouped under the generic name of tauopathies, including approximately one-half of all frontotemporal dementia (FTD) cases and Alzheimer's disease (Boeve, 2012). Both rare (Coppola et al., 2012; Hutton et al., 1998) and common (Hoglinger et al., 2011) genetic variation have been shown to mediate risk for tauopathies. The major common variant risk for PSP, a prototypical tauopathy, involves a region surrounding the tau locus (Conrad et al., 1997), but how such genetic variation might mediate risk is not known.

We profiled the methylation status in peripheral blood from patients with two tau-related neurodegenerative conditions, PSP and FTD, using Illumina DNA methylation arrays. We then

integrated these methylation data with SNP and gene expression data to identify a mediating role for methylation in genetic risk for PSP. We replicate this finding in independent studies and show that it is conserved in brain, providing the first evidence for a role for DNA methylation in mediating the risk for neurodegenerative dementia.

Sirtuin 1 (SIRT1), a member of the sirtuin family, plays an important role in key cellular processes, including senescence/aging and inflammation (Gan and Mucke, 2008; Libert and Guarente, 2013). SIRT1 deacetylates intracellular targets, including transcription factors, signaling molecules, and histones (Michan and Sinclair, 2007; Zhang et al., 2011). In a mouse model, Cho et al. demonstrated that SIRT1 deficiency contributes to the selective activation of IL-1β transcription through hypomethylation of the specific CpG sites on the IL-1β proximal promoter (Cho et al., 2015). We further apply these data to demonstrate an association between hypomethylation of the IL-1β promoter and IL-1β mRNA expression, aging, and neurodegenerative disease.

Differential methylation analysis

We first analyzed methylation profiles in 171 patients with FTD (n=128) and PSP (n=43) and compared them with 185 subjects with no evidence of dementia or other neurological conditions using Illumina HumanMethylation 450k arrays (Supplementary Table B-1). Two datasets were generated in two batches, samples were compared within each dataset to condition out a potential batch effect, and the resulting differentially methylated probes (DMPs) were combined (see Methods).

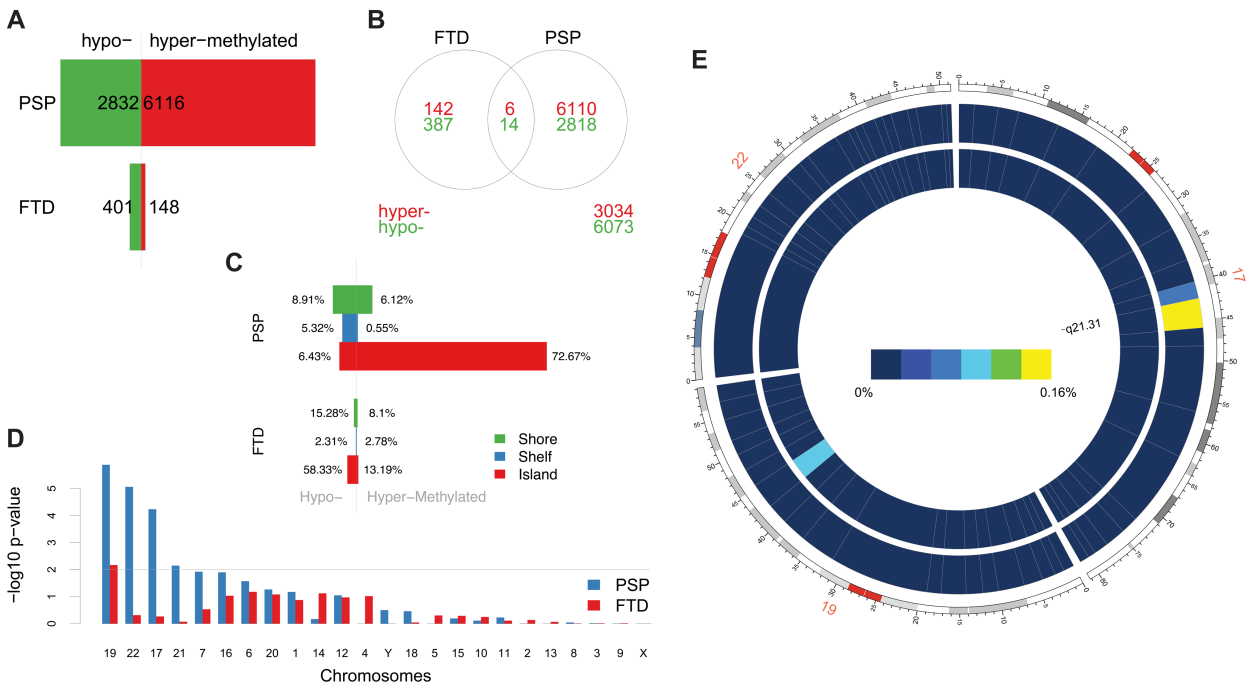


Figure 5-1: (a) Barplots representing the numbers of differentially methylated probes (DMPs) identified in each disease group vs. controls (Benjamini-Hochberg-adjusted *p*-value ≤0.05). The number of DMPs indicated in PSP vs. Control comparison is the union set of DMPs identified in dataset #1 and dataset #2. Red bars: hypermethylated DMPs, green bars: hypomethylated DMPs. **(b)** Venn diagram representing the overlap between DMPs in FTD vs. controls and PSP vs. controls. Red numbers: hypermethylated DMPs; green: hypomethylated DMPs. **(c)** Barplots representing DMPs classified by probe type. CpG island probes are overrepresented in both FTD vs. controls and PSP vs. controls. **(d)** Chromosome enrichment analysis: DMPs are significantly enriched in chromosomes 19, 22, and 17, only in PSP vs. controls (y axis: −log10 (p-value), hypergeometric test). **(e)** Circos plot of chromosomes 19, 22, and 17 showing regional enrichment of DMPs (PSP vs. Control comparison, BH adjusted p-value ≤0.05, absolute average beta difference (aβD)>0.1) in one region on chromosome 17. Each chromosome was divided into 20 regions, which contain the equal number of CpG probes. Regions were colored according to the DMP density. Blue: low DMP density, yellow: high density. Circles from inner to outer represent FTD, PSP vs. controls, respectively.

Differential methylation analysis identified a number of DMPs between affected subjects and controls, with partial overlap between PSP and FTD (Figure 5-1a-b, complete list of DMPs is in Supplementary Table C-2). DMPs were mostly clustered within CpG islands (defined according to the Illumina annotation), with most being hypermethylated in PSP vs. controls (Figure 5-1c, Table 5-1). Gene ontology analysis of DMPs in PSP vs. controls showed overrepresentation of genes involved in a number of pathways, including DNA binding and transcription factor binding (Supplementary Figure C-1). We then assessed the chromosomal distribution of the DMPs, and observed – only in PSP samples vs. controls – an overrepresentation of probes from chromosomes 19 (hypergeometric test *p*-value = 1.32×10^{-6}), 22 ($p = 8.63 \times 10^{-6}$), and 17 (p =

Table 5-1: DMPs identified in disease vs. controls classified by probe type (Island, Shelf, and Shore).

	Methylation status	Island	Shelf	Shore	p-value	overall p-value
FTD	Hyper	57	12	35	0.709	2.32×10^{-17}
	Hypo	252	10	66	1.18×10^{-20}	
PSP	Hyper	5413	41	456	0	0
	Hypo	479	396	664	2.12×10^{-70}	

Hypomethylated DMPs are overrepresented in CpG islands when comparing FTD vs. controls, and both hyper- and hypomethylated DMPs are overrepresented in CpG islands in PSP vs. controls. Chi-square test was performed within hyper and hypo-methylated DMPs (p-value column) or across all DMPs (overall p-value column) by considering the proportion of all probes on the chip as population frequencies.
doi:10.1371/journal.pgen.1004211.t001

5.82 × 10^{-5}, Figure 5-1d), with most top DMPs (after filtering for absolute average beta difference (abD) >0.1) located within the 17q21.31 region (Figure 5-1e). The most significant DMPs when comparing PSP vs. controls (n = 14, absolute abD >0.1) are listed in Table 5-2. Of note, 4 DMPs (all hypomethylated in PSP) were located within the *NFYA* gene, encoding for a component of a nuclear transcription factor. Importantly, 3 of the 14 significant DMPs are located in 17q21.31 (Figure 5-1e, $p = 2.23 \times 10^{-7}$, hypergeometric test). Despite being located in a relatively limited genomic region, these 3 probes were both hypermethylated and hypomethylated in PSP vs. controls, suggesting complex disease-associated patterns of differential methylation in this region.

17q21.31 haplotype and methylation

The location of several DMPs in the 17q21.31 region was intriguing because the 17q21.31 locus contains an established risk factor for neurodegeneration, first reported in 1997 by Conrad *et al.* (Conrad et al., 1997) for PSP and then confirmed in multiple series (Wade-Martins, 2012). Two main haplotypes (H1 and H2) have been described at this locus. The more common H1 haplotype is over-represented (95% vs. 57%) in PSP vs. normal controls (Conrad et al., 1997;

Table 5-2: Top DMPs identified in PSP vs. controls, after filtering for an adjusted p-value ≤0.05, and an absolute average beta difference (aβD)≥0.1.

Probe ID	aβD	aMeth	P-value	Adjusted P-value	Gene	Position	Probe Type
cg03865648	−0.105	0.286	3.37×10^{-5}	0.007	/	chr3:173113856	N_Shore
cg09580153	−0.100	0.920	4.55×10^{-5}	0.008	NFYA	chr6:41068724	Island
cg12000995	−0.109	0.535	5.23×10^{-5}	0.009	KRTCAP3	chr2:27665139	Island
cg03644281	−0.105	0.916	8.78×10^{-5}	0.012	NFYA	chr6:41068752	Island
cg04346459	−0.131	0.876	9.03×10^{-5}	0.012	NFYA	chr6:41068666	Island
cg03428951	−0.133	0.588	1.16×10^{-4}	0.014	FAM153C	chr5:177434336	S_Shore
cg25110423	−0.112	0.848	1.23×10^{-4}	0.015	NFYA	chr6:41068646	Island
cg23758822	**0.105**	**0.205**	$\mathbf{1.74\times10^{-4}}$	**0.018**	**/**	**chr17:41437982**	**N_Shore**
cg22968622	**−0.127**	**0.160**	$\mathbf{1.85\times10^{-4}}$	**0.019**	**/**	**chr17:43663579**	**Island**
cg22295435	0.140	0.564	1.95×10^{-4}	0.019	VSTM2A	chr7:54615864	S_Shore
cg21819782	−0.133	0.294	3.89×10^{-4}	0.028	/	chr2:62609317	NA
cg12609785	**−0.103**	**0.129**	$\mathbf{4.10\times10^{-4}}$	**0.029**	**/**	**chr17:43660871**	**N_Shore**
cg24401049	−0.102	0.578	7.21×10^{-4}	0.040	ARHGAP6	chrX:11157158	Island
cg12289251	−0.106	0.389	8.08×10^{-4}	0.043	CACNB2	chr10:18689471	NA

ID: Illumina probe ID; aβD: average beta difference, aMeth: average Methylation level. In bold are probes located within the 17q21.31 region.
doi:10.1371/journal.pgen.1004211.t002

Kalinderi et al., 2009; Wade-Martins, 2012). The H1/H2 locus spans at least 1.8 Mb and includes multiple genes (>40, many of which are actively transcribed in the brain), notably including *MAPT*, encoding for the microtubule-associated protein tau (Wade-Martins, 2012). Mutations in *MAPT* cause FTD and PSP, and hyperphosphorylated tau accumulation is a hallmark in a number of neurodegenerative conditions, including AD, PSP, FTD and others, collectively named 'tauopathies'.

Consistent with previous reports, the H1 haplotype was overrepresented in our PSP cohort, with a H1 allelic frequency of 97.1% vs. 80.4% in controls (p = 1.86 × 10^{-4}, Fisher's exact test, Supplementary Table C-1), further confirming – even in this relatively small data set – the H1 haplotype as a risk factor for PSP. We hypothesized that the clustering of DMPs in 17q21.31 in PSP cases vs. controls might be related to the H1 haplotype risk factor. To detect an effect of the 17q21.31 haplotype on methylation levels, we compared samples based on their genotype at this region, independent of disease classification. As for previous analyses, we compared

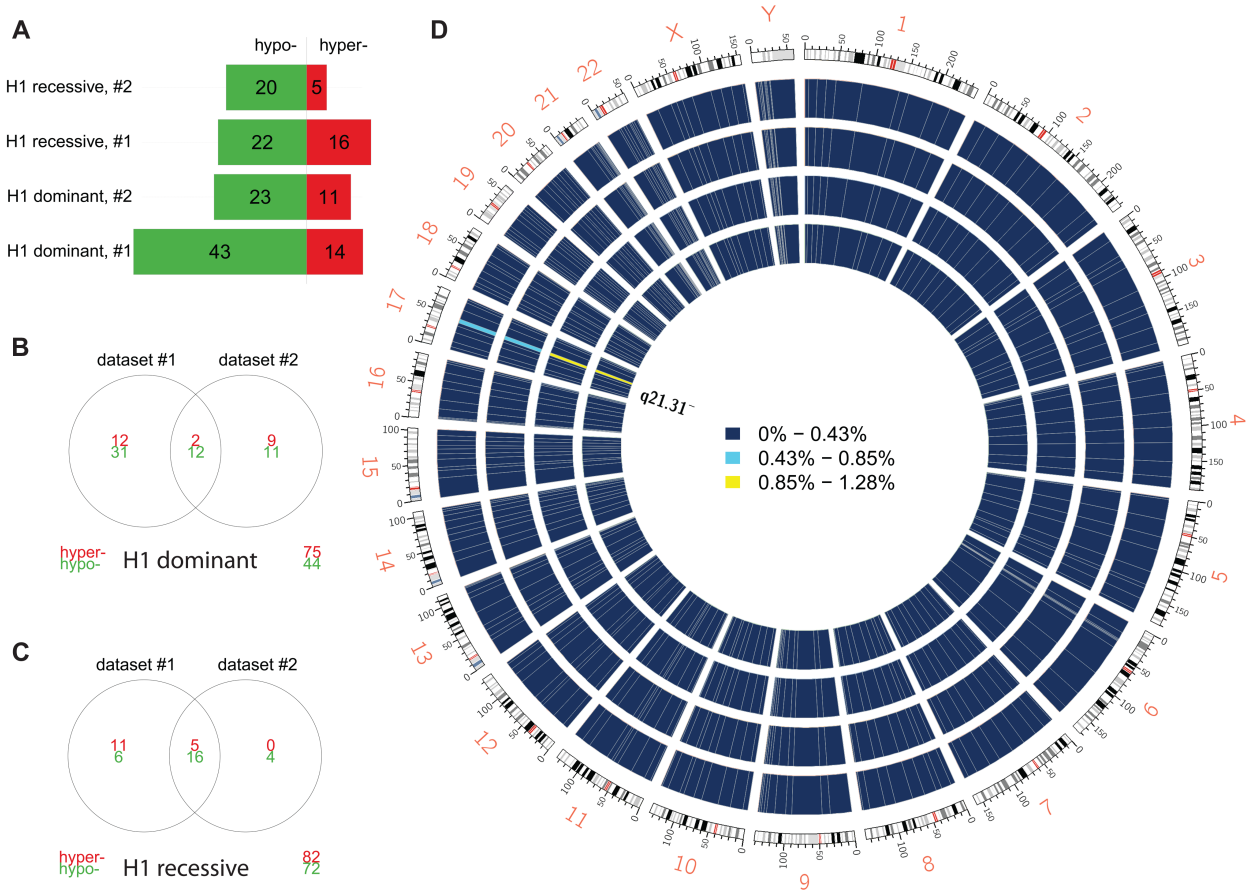


Figure 5-2: (a) Number of DMPs (Benjamini-Hochberg-adjusted *p*-value ≤0.05) identified in each comparison and each dataset. Dominant: dominant model (H1H1+H1/H2 vs. H2/H2); recessive: recessive model (H1H1 vs. H1/H2+H2/H2). **(b)** Overlap between datasets #1 and #2 (dominant model). **(c)** Overlap between datasets (recessive model). **(d)** Circos plot showing the physical density across the genome of DMPs. Each chromosome was divided in 10 regions, and the proportion of DMPs was assessed. Regions were colored according to the DMP density. Blue: low DMP density, yellow: high density. Circles from inner to outer represent Dataset #2, recessive model; Dataset #1, recessive model; Dataset #2, dominant model; Dataset #1, dominant model. DMPs were mostly enriched in chr17q21.31.

samples within datasets to avoid potential batch effects. Genotype distribution across diseases and datasets is reported in Supplementary Table C-2.

We compared carriers of the risk-associated H1 haplotype (H1/H1 and H1/H2 genotypes) to H2/H2 samples (dominant model) within each dataset and, after filtering DMPs for adjusted $p \leq 0.05$, identified two overlapping sets of 57 and 34 DMPs (Figure 5-2a,b), markedly clustered within the 17q21.31 region (Figure 5-2d). Similar results (Figure 5-2a,c) were obtained when comparing H1/H1 samples to H2 carriers (H1/H2 and H2/H2, recessive model), supporting the

Table 5-3: DMPs identified when comparing 17q21.31 H1 carriers to non-carriers (dominant model, absolute average beta difference (aβD)>0.1, adjusted p-value ≤0.05).

	Dataset #1		Dataset #2				
ID	**aβD**	**adjusted p**	**aβD**	**adjusted p**	**Gene**	**Coordinate**	**Type**
cg18878992	−0.185	3.71×10^{-55}	−0.341	1.81×10^{-86}	MAPT	chr17:43974344	Island
cg17117718	−0.263	7.00×10^{-28}	−0.212	3.90×10^{-19}	/	chr17:43663208	Island
cg07870213	−0.203	3.50×10^{-19}	−0.174	2.82×10^{-23}	DND1	chr5:140052090	Island
cg16228356	0.167	8.74×10^{-12}	0.145	2.70×10^{-04}	/	chr17:43848958	/
cg04703951	−0.186	5.70×10^{-08}	−0.158	2.04×10^{-07}	/	chr17:43578652	/
cg23955979	−0.215	9.14×10^{-06}	−0.185	3.19×10^{-12}	/	chr17:45126661	/
cg00846647	0.133	6.95×10^{-05}	0.178	6.62×10^{-19}	MAPT	chr17:44060252	Island
cg19832721	−0.254	8.59×10^{-03}	−0.295	1.73×10^{-05}	KIAA1267	chr17:44249866	/
cg22968622	−0.342	2.26×10^{-02}	−0.369	2.04×10^{-07}	/	chr17:43663579	Island

doi:10.1371/journal.pgen.1004211.t003

hypothesis of a strong *cis* effect of the H1/H2 locus on methylation levels in peripheral blood (Figure 5-2). After filtering for absolute aβD >0.1, 8 of the top 9 DMPs identified in both datasets were within 17q21.31 (Table 5-3) in the dominant model. As noted in PSP cases vs. controls, DMPs in this region are both hyper- and hypo-methylated, suggesting a complex cis-regulation of methylation levels (Figure 5-3a). Scatterplots of the methylation levels for the top DMPs shared between the dominant and recessive models indicate that the H1 haplotype influences methylation levels at these sites in a dose-dependent fashion (Figure 5-3b), accounting for a majority of methylation variability at these sites (e.g. R-squared = 0.835 and 0.866 in dataset #1 and #2, respectively, for cg22968622). Similar results were obtained when comparing subjects based on their genotype at 17q21.31, but only within controls, FTD, or AD patients (Supplementary Material). The H1 haplotype can be further divided into sub-haplotypes (Pittman et al., 2005). We obtained sub-haplotype information for 93 H1 carriers in our cohort using the SNPs described in Kauwe et al. (2008). Hierarchical clustering of the methylation signal in the 17q21.31 region and principal component analysis did not reveal a particular clustering of H1 sub-haplotypes (data not shown). These results – although based on a subset of our cohort – suggest that haplotype structure is not the major determinant of 17q21.31 methylation overall.

To test the contribution of haplotype status on PSP-associated DMPs, we repeated the differential methylation analysis only on samples with the H1/H1 haplotype (n=31 PSP cases

and 59 unaffected controls). Of the resulting 341 significant DMPs (after application of the

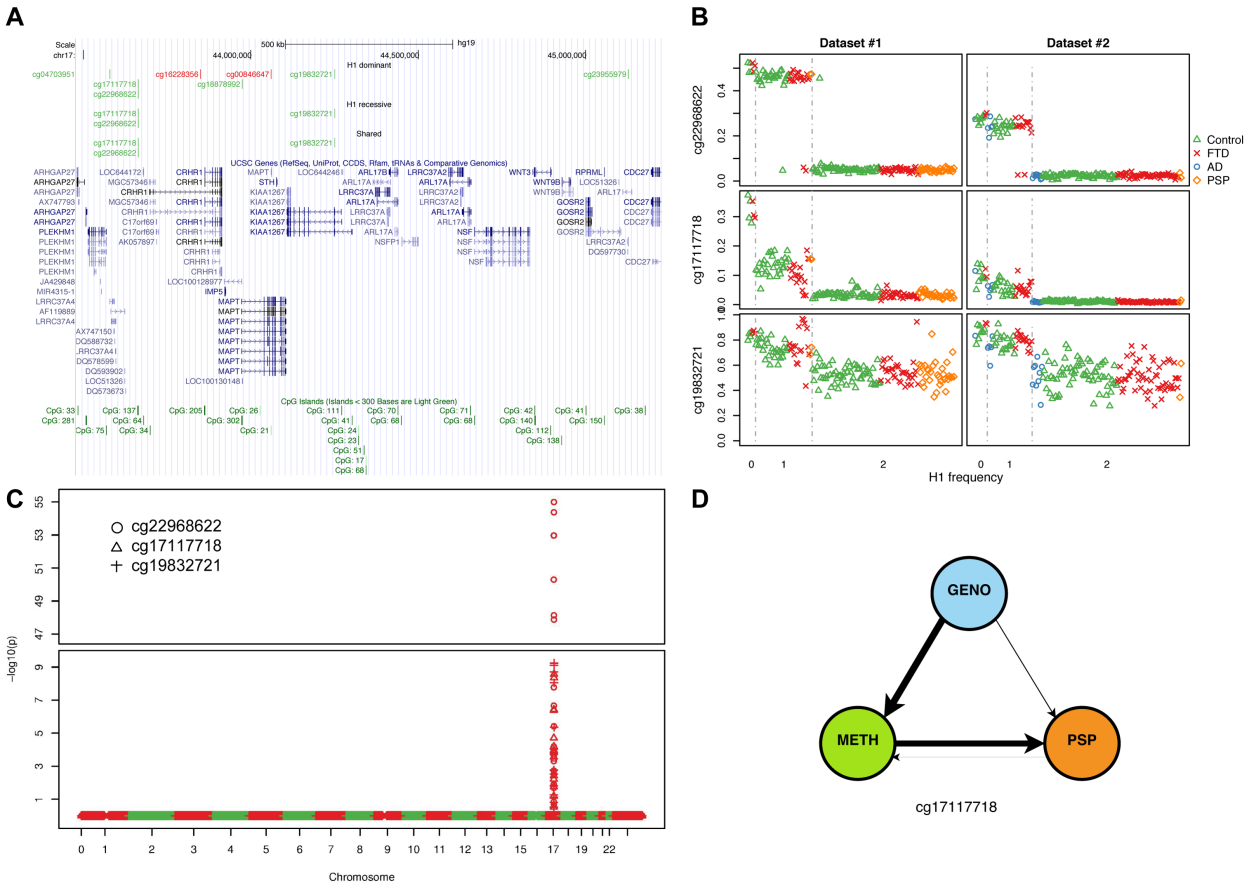


Figure 5-3: Methylation-QTL at 17q21.31.
(a) Physical position of top (BH adjusted p-value ≤0.05, absolute average beta difference (aβD)>0.1) DMPs identified when comparing samples based on 17q21.31 haplotype. Dominant: dominant model (H1H1+H1/H2 vs. H2/H2); recessive: recessive model (H1H1 vs. H1/H2+H2/H2); Shared: DMPs shared between the two previous comparisons. Red: hypermethylated, Green: hypomethylated. **(b)** Scatterplot of the methylation levels of 3 top DMPs identified from both H1 dominant and recessive model. **(c)** Methylation-QTL analysis performed in 226 individuals of European descent on 3 the top DMPs identified when comparing H1 vs. H2 haplotypes. Manhattan plot representing p-values by chromosome. At each genomic location the smaller −log10 p-value from two datasets was plotted. A single cluster at 17q21.31 was identified for all three DMPs. **(d)** Results of network edge orienting (NEO) analysis for the differentially methylated probe cg17117718 following the mediation model (GENO causes METH causes PSP). The arrow line thickness is proportional to the likelihood that the edge is oriented in the causal direction, found by calculating the relative probability of the model likelihoods determined by NEO.

Benjamini-Hochberg procedure, FDR = 0.05), 21 were located in chromosome 17 and 2 were

located in the 17q21.31 band. Neither the chromosome nor the region were found to be significantly overrepresented by the hypergeometric test (p = 0.342 and 0.149, respectively). The lack of overrepresentation within 17q21.31 after conditioning on strata defined by the 17q21.31 haplotype suggests that either the previously identified overrepresentation on chromosome 17 was due to the 17q21.31 haplotype effect on methylation levels, or that it could also reflect reduced power due to small sample size. To address the issue that the strata contained too few samples, we also carried out a multivariate regression model analysis that included 17q21.31 haplotype as covariate. Specifically, the methylation level of each of the 3 top PSP-related DMPs located in 17q21.31 (Table 5-2) was regressed on PSP status, 17q21.31 haplotype, ethnicity, and age using a multivariate linear regression model. We found that, except for cg23758822, other PSP-related DMPs were no longer significant (p = 0.410 on average, Supplementary Table C-3) in a multivariate model once it included the H1 genotype. We also calculated the relative weight of each predictor using the R package relaimpo (Groemping, 2006) and determined that the H1 haplotype accounted for the majority of explained variance (78.2 ± 25.9%, Supplementary Figure C-2). Finally, we estimated relative cell count composition in peripheral blood using methylation data (Houseman et al., 2012; Koestler et al., 2013; Liu et al., 2013). Correction for inferred cell count did not significantly change our findings (Appendix C, Supplementary Table C-4, Supplementary Figure C-11).

Taken together, these findings indicate 1) a strong effect of the 17q21.31 haplotype on methylation levels at 17q21.31, 2) that the risk-associated H1 determines most of the methylation changes observed with confidence in PSP patients vs. controls, and 3) that additional DMPs outside the 17q21.31 region may be at play in determining risk susceptibility for PSP in H1 carriers, though larger sample sizes will be needed to clarify their importance.

Genome-wide methylation QTL analysis confirms a *cis* methQTL at 17q21.31

Our findings strongly indicate a *cis* regulation of methylation levels at the 17q21.31 locus. To test whether there were additional potential genetic determinants of methylation levels at 17q21.31 in our dataset, we performed a methylation QTL (methQTL) analysis in a subset of 226 individuals of European descent for whom whole-genome SNP and methylation data were available (Supplementary Table C-5). We assessed association of genetic variants with methylation levels at 3 CpGs within 17q21.31 (cg22968622, cg17117718, cg19832721) in each dataset. We identified on average 110 genome-wide significant signals (Bonferroni-adjusted $p \leq 0.05$), all located within the 17q21.31 region (Figure 5-3c, Supplementary Table C-6). These variants accounted for a proportion of variability ranging between 25.5% and 98.2% (mean R-squared = 0.701, Supplementary Figure C-3, Supplementary Table C-7) further confirming that genetic variants at 17q21.31 are controlling methylation levels in *cis* in the same region. We focused on Caucasian individuals because of the differences in frequency of the H2 haplotype across populations. In fact, consistent with previous reports (Evans et al., 2004), we observed that the H2 haplotype occurs more frequently in Caucasians (H2 allelic frequency = 19.2%, Table 4) than in other ethnic groups (H2 allelic frequency in Asians = 1.3%; $p = 3.83 \times 10^{-6}$, Fisher's exact test). However, similar results were observed when including all the 273 individuals for whom SNP and methylation data were available (Appendix C, Supplementary Figure C-4, Supplementary Table C-6, Supplementary Table C-7).

***Cis* methQTL effects at the 17q21.31 locus in additional datasets**

To confirm that the 17q21.31 haplotype regulates methylation in *cis* at this locus in an independent dataset, we downloaded and reanalyzed raw data from a previously published study, for which SNP and methylation data in peripheral blood from 12 samples were publicly available (Heyn et al., 2012). Using the rs1052553 SNP to call the H1/H2 haplotype and adopting the same statistical thresholds, we compared H1/H1 vs. H1/H2 subjects and identified

Table 5-4: Relative distribution of haplotypes at 17q21.31 in ethnic groups.

	Dataset #1			Dataset #2		
	H2H2	**H1H2**	**H1H1**	**H2H2**	**H1H2**	**H1H1**
Caucasian	6	42	90	9	33	94
Asian	0	1	18	0	0	21
Latino	1	5	11	2	4	13
Unknown	0	1	4	0	2	2
Total	**7**	**49**	**123**	**11**	**39**	**130**

Ethnicity was inferred for 271 samples using SNP clustering compared to Hapmap data (see Methods). Self-reported ethnicity was used for an additional 88 samples.

doi:10.1371/journal.pgen.1004211.t004

one hypomethylated probe (cg22968622, adjusted *p*-value = 2.37×10^{-8}, aβD =-0.42) within 17q21.31, which was also identified in our analysis. We also performed a methQTL analysis for cg22968622 in the same dataset. Of the 310 significant SNPs, 206 were located in the 17q21.31 region ($p = 0$, hypergeometric test), further supporting the presence of a *cis* methQTL at this locus.

To provide independent validation of the methylation array assay, we performed reduced representation bisulfite sequencing (RRBS) on a representative set of 7 samples from the study (2 H1/H1 controls, 1 H1/H1 PSP patient, 1 H1/H2 control, 1 H1/H2 PSP patient, and 2 H2/H2 controls). As a sequencing-based approach, RRBS would not suffer from some of the technical biases present in arrays, e.g. due to hybridization. At CpG sites that were covered by both RRBS and array, the methylation measurements were highly correlated (Pearson $r > 0.9$) in all seven samples (Supplementary Figure C-5).

To validate our findings from peripheral blood, we analyzed RRBS data from whole-blood DNA of a separate cohort of 80 healthy subjects (comprising 54 H1/H1, 24 H1/H2, and 2 H2/H2). On average, the methylation level computed from RRBS was highly correlated with the array in both dataset #1 (r = 0.965) and dataset #2 (r = 0.963) (Supplementary Figure C-6). Consistent with

the array, we found differences in methylation that were significant even after strict Bonferroni correction for multiple testing, mostly localized to the 17q21.31 cytoband (Supplementary Table C-8). Since local methylation levels are often highly correlated, we would expect the differentially methylated CpGs identified by both the array and the sequencing method to be in close proximity. Indeed, the differentially methylated loci identified by RRBS in the 17q21.31 region are nearby those identified by the Illumina Human Methylation array, overlapping the same genes MAPT and KIAA1267. This degree of overlap is striking, given the large extent of the haplotype inversion (Supplementary Figure C-7). Methylated regions identified by the array but not by RRBS may be a result of the higher power (greater sample numbers) in the array, and differences in coverage. Therefore, we looked at probes that both demonstrated haplotype-specific methylation in 17q21.31 on the Illumina array and were covered by RRBS reads in the additional cohort. One probe, cg08113562, met these criteria. The methylation pattern followed a statistically significant dose-dependent relationship with the H1 versus H2 haplotype, with mean methylation fractions of 0.001 in H1/H1 subjects, 0.022 in H1/H2 subjects, and 0.048 in H2/H2 subjects (two-sided p = 0.03, ANOVA), in the same direction as that reported by the array.

To assess the relevance of our findings to brain tissue, we analyzed 2 published methylation QTL studies in brain involving 150 (Gibbs et al., 2010) and 153 (Zhang et al., 2010) subjects, respectively. Gibbs *et al.* (Gibbs et al., 2010) identified 9 SNP-CpG association pairs (out of 52,345 significant methQTL, mean R-squared = 0.232) at the 17q21.31 locus in two (frontal cortex and cerebellum) of the four studied brain regions; Zhang *et al.* (Zhang et al., 2010) identified in cerebellar samples 122 SNP-CpG pairs (significant in at least one of the three thresholds they used) at the 17q21.31 locus (out of 12,117 significant methQTLs, mean R-squared = 0.136). Together, these data demonstrate that the *cis* methQTL we identified in our study are present in independent studies in peripheral blood, and are preserved in brain.

Causal inference identifies three methylated regions that may mediate PSP risk

The strong association between the haplotype at 17q21.31 and methylation status raises the question of whether methylation levels mediate the protective or pathogenic effects of haplotype variants. Recent developments in the field of causal inference have yielded quantitative methods to predict the hierarchy of causation given genetic variants (Pearl, 2009; Schadt et al., 2005; Vansteelandt and Lange, 2012; Zhu et al., 2007). Network Edge Orienting (NEO), for example, uses structural equation models to choose the best fitting causal model, assuming that the genetic variation is fixed by meiosis and thus "anchors" each model (that is, genotype precedes phenotype) (Aten et al., 2008). NEO allows one to evaluate which of five testable causal models (Supplementary Figure C-8) best explains the relationship between genetic variants, methylation levels, and disease status. For instance, the genotype may lead to patterns of methylation that directly contribute to the disease phenotype (Supplementary Figure C-8b). Under this model, the DMPs within 17q21.31 would be the most interesting, as they would correspond to the epigenetic markers mediating the increased risk conferred by the H1 haplotype. Alternatively, the genotype may independently give rise to the methylation and disease phenotype, with neither contributing to the other (Supplementary Figure C-8c). DMPs under this model are only associated with the disease because of the common source of variation due to the 17q21.31 locus.

We applied NEO to calculate a relative fitting index of the "mediation model" for the 9 haplotype-associated DMPs (Supplementary Table C-9) using 35 PSP cases and 184 unaffected controls for whom these data were available. The "mediation model" best explained the methylation pattern in three sites, one of which (cg17117718) was statistically significant (see Methods) (Figure 5-3d). These results support the hypothesis that methylation status at certain sites likely is a causal mediator of the major known genetic risk related to PSP pathogenesis. Taken

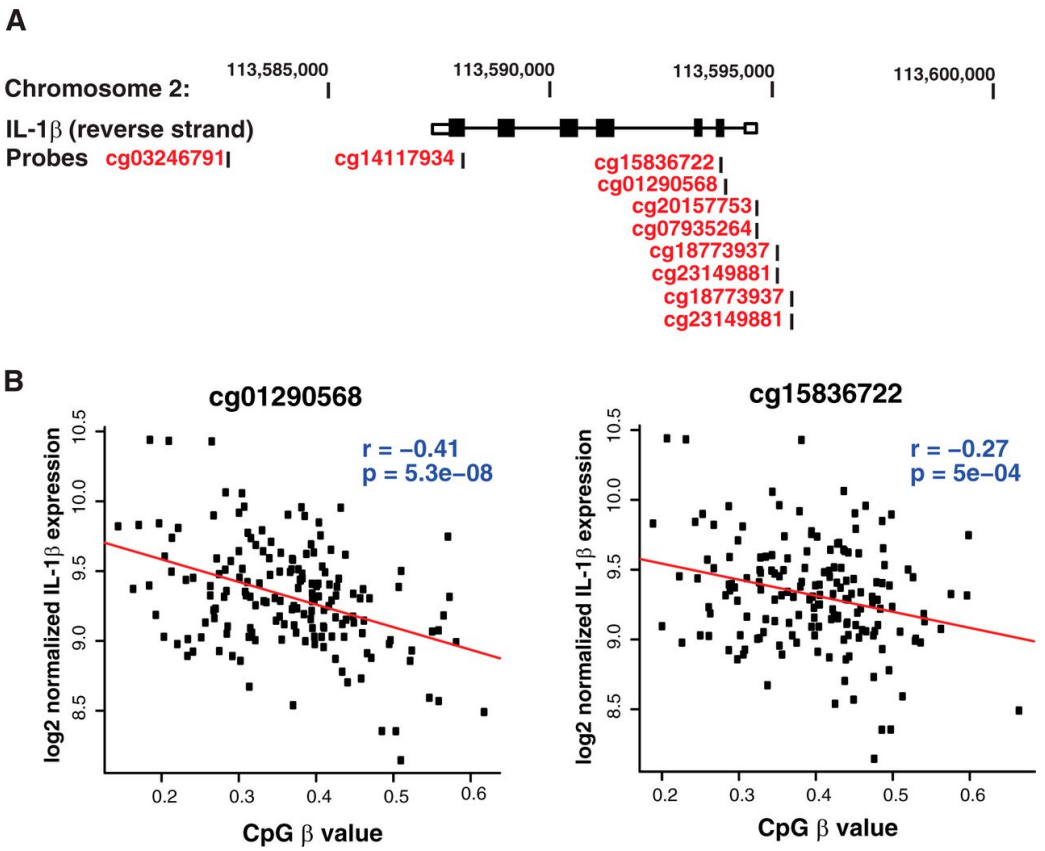


Figure 5-4: IL-1β CpG hypomethylation is associated with increased IL-1β expression. **(a)** Schematic diagram of the human IL-1β gene on chromosome 2 and 10 CpG sites covered by the Illumina HumanMethylation 450K array. Coordinates on chromosome 2 are based on hg19/GRCh37. Boxes represent exons, whereas the connecting lines represent introns; filled boxes correspond to coding sequence, and unfilled boxes correspond to UTR (untranslated region). **(b)**, Illumina HumanMethylation 450K was used to assay methylation of DNA from peripheral whole blood. mRNA expression in peripheral whole blood was quantified with the Illumina HumanHT-12 v4.0 Gene Expression BeadChip. Methylation levels at cg01290568 and cg15836722 negatively correlated with expression of IL-1β in 167 whole-blood samples.

together, these results predict – for the first time – a link between epigenetic changes and tauopathies, and will need to be further validated with functional studies.

Haplotype-associated differences in gene expression

Methylation changes have been associated with changes in gene expression. We examined microarray expression data in peripheral blood available for 120 subjects, to test whether the methylation associated with the 17q21.31 haplotype had such an effect. Among 88 healthy subjects with H1H1 haplotype, 24 healthy subjects with the H1H2 haplotype, and 8 healthy subjects with the H2H2 haplotype, we identified three probes significantly differentially

expressed in peripheral blood, mapping to *MAPK8IP1* (on chromosome 11), *LRRC37A4* (located within the 17q21.31 region), and *MTFP1* (on chromosome 22, Benjamini-Hochberg adjusted *p*-values of 2.4×10^{-20}, 8.4×10^{-5}, and 4.0×10^{-2}, respectively). The three probes demonstrated a log fold change of 0.69, -0.25, and 0.10, respectively, in H2 vs. H1 carriers.

The strong haplotype-associated methylation changes identified in our study included DMPs within the *MAPT*, *KIAA1267*, *ARHGAP27*, and *DND1* genes. We used linear regression to test whether the 17q21.31 haplotype was associated with differential expression of these genes, and found no correlation between haplotype and gene expression for these transcripts (adjusted R-squared = 0.006, 0.000, -0.008, and -0.011, respectively). Thus, while haplotype was shown to affect mRNA expression of *MAPK8IP1* and *LRRC37A4* in peripheral blood, there was no detectable correlation between DMP-containing genes and their corresponding expression levels.

IL-1β CpG hypomethylation is associated with increased IL-1β expression

To validate the relationship between IL-1β hypomethylation with IL-1β expression in humans, we used the Illumina HumanMethylation 450K to assay methylation of human blood samples from 167 subjects, in which levels of IL-1β mRNA were obtained using Illumina HumanHT-12 v4.0 Gene Expression BeadChip. Ten probes were found near the IL-1β gene (a). Consistent with the finding that hypomethylation of IL-1β upregulates its expression in mouse myeloid cells, methylation of two CpGs (cg01290568 and cg15836722) negatively correlated with IL-1β expression in humans (b). This suggests that, in humans, IL-1β transcript levels are regulated by methylation.

Two IL-1β CpG sites are hypomethylated in normal aging and in demented patients with tauopathy

Because microglial SIRT1 was decreased with the aging of microglia and lack of SIRT1 decreased DNA methylation in myeloid cells (Cho et al., 2015), we hypothesize that inverse relationship exists between IL-1β methylation and aging in humans. The Illumina HumanMethylation 450K was used to assay methylation in three batches of human blood samples. Methylation of cg01290568 and cg15836722 was inversely correlated with chronological age, after correction for multiple comparisons (Figure 5-5a). To account for a possible batch effect, samples were colored by array batch, and no confounding effect was observed. We compared IL-1β methylation in dementia patients and nondemented controls. To exclude the possible confounding effect of chronological age in this analysis, the residual of the methylation level after regression for age was used. Two CpG sites (cg01290568 and cg15836722) that were hypomethylated with normal aging were also hypomethylated in demented patients (Figure 5-5b). The hypomethylation appears to be highly selective as no other CpG sites were affected by aging or tau-mediated dementia (data not shown).

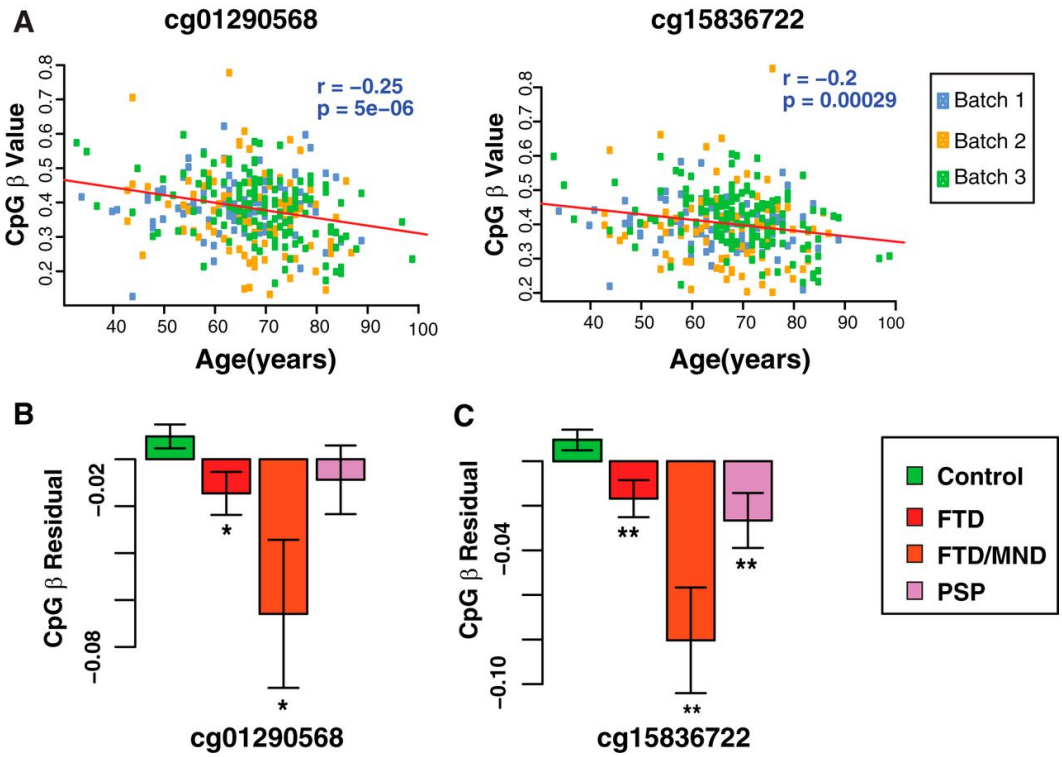


Figure 5-5: Two IL-1β CpG sites are hypomethylated in normal aging and in demented patients with tauopathy. **(a)**, Methylation at cg01290568 and cg15836722 correlated with chronological age of 335 nondemented controls assayed by Illumina HumanMethylation 450K array. Data for each CpG were fitted to a linear model (red line). Samples were colored according to array batch. **(b, c)**, Average methylation residual at cg01290568 or cg15836722 in nondemented controls (*n* = 335) and patients with FTD (*n* = 122), FTD/MND (*n* = 9), and PSP (*n* = 43). Two-sided *t* test was used to compare dementia patients versus nondemented controls. Values are mean ± SEM.

Discussion

The goal of this study was to assess whether changes in DNA methylation in peripheral blood are observed in patients with neurodegenerative diseases. By performing microarray-based differential methylation analysis, we identified a methylation signature associated with disease status in PSP and, to a lesser extent, FTD. Using SNP data available in a subset of our series, we showed that a remarkable proportion of the observed changes in methylation status in PSP are associated with a common haplotype at the 17q21.31 locus, strongly suggesting the presence of a *cis* methylation QTL in this region. Although we included patients with neurodegenerative disorders in our analysis, the observed pattern seems to be related to the haplotype at 17q21.31, independent of disease status. Integrative analyses including SNP and

gene expression data support a model whereby genetic variation at the 17q21.31 locus modulates the risk for neurodegenerative tauopathy at least partially via differential methylation.

The H1 haplotype at 17q21.31, a large linkage disequilibrium block due to an inverted chromosomal sequence of ~970 kb, is the major known risk locus for PSP (Baker et al., 1999) and other neurodegenerative diseases. Although the genetic contribution of this locus to the risk for neurodegeneration is established and widely replicated, the mechanism by which risk is increased is largely unknown. This region spans from *CRHR1* (corticotrophin) to *IMP5* (a presenilin homologue) at the centromeric end of LD, while *WNT3* and *NSF* (N-ethylmaleimide-sensitive factor) are at the telomeric end of the LD block (Pittman et al., 2004); therefore it spans at least 1.8 Mb, including 48 RefSeq genes – many of which actively transcribed in the brain – and constitutes the largest haplotype block in the human genome. Stefansson et al. (2005) showed that the complete disequilibrium was due to an inversion occurring in the H2 haplotype relative to the H1 human reference and subsequent absence of recombination between inverted and non-inverted chromosomes. The study of this region to understand susceptibility to neurodegeneration has been mostly focused on one gene, *MAPT*, encoding for the microtubule-associated protein tau. This focus on tau is well motivated, as hyperphosphorylated tau accumulates within neurofibrillary tangles – the pathological hallmark of AD – and because mutations in *MAPT* cause FTD, the second most common neurodegenerative dementia. Several *in vitro* studies have reported alterations of transcription levels in *MAPT* due to common variants in the region (Myers et al., 2007; Rademakers et al., 2005), but this finding has not been consistently replicated (Hayesmoore et al., 2009; Trabzuni et al., 2012). More consistent evidence exists for a higher expression of exon 3 in brains from H2 carriers (Caffrey et al., 2008; Trabzuni et al., 2012) and of exon 10 in H1 carriers (Caffrey et al., 2006; Myers et al., 2007), suggesting that splicing abnormalities are involved in increasing risk.

Recently, additional genetic evidence has been reported implicating this locus in other neurodegenerative diseases, such as Parkinson's disease (Simon-Sanchez et al., 2009), essential tremor, and multisystem atrophy (Vilariño-Güell et al., 2011). This is important, as these diseases are not typical tauopathies, suggesting that the effect of this risk-associated region may be complex and involve multiple genes (Hoglinger et al., 2011).

Our results indicate a novel mechanism by which the H1/H2 locus may affect the risk for tauopathies: significant alterations in methylation mediating increased disease susceptibility. Importantly, these methylation changes are not at the *MAPT* locus only, but are consistently observed in at least 3 neighboring genes as well, suggesting that genes other than *MAPT* might be at play in increasing disease susceptibility. In addition, DMPs in the region were both hyper and hypo-methylated, suggesting a complex regulation of methylation levels at this locus. Further studies will be needed to understand whether the observed methylation signature at 17q21.31 is increasing susceptibility through a *MAPT*-dependent or independent mechanism, or both. Interestingly, a recent study focused on rheumatoid arthritis (Liu et al., 2013) linked a genetic susceptibility region for the disease, the MHC locus, with methylation changes in the same region, supporting the notion that epigenetic changes might mediate complex disease susceptibility induced by genetic risk factors.

This is the first study of DNA methylation levels in blood in PSP and FTD, disorders that mainly affect brain. Although methylation patterns may be tissue specific (Ghosh et al., 2010), comparative studies of blood and brain showed both methylation patterns that are tissue-specific and conserved across tissues (Davies et al., 2012). We show that the particular H1 haplotype-related methylation pattern identified in blood is at least partially conserved in brain. This is encouraging, since – in contrast to brain – blood is available from living patients, yielding a higher potential for future use as biomarker and the possibility of large-scale studies. We and

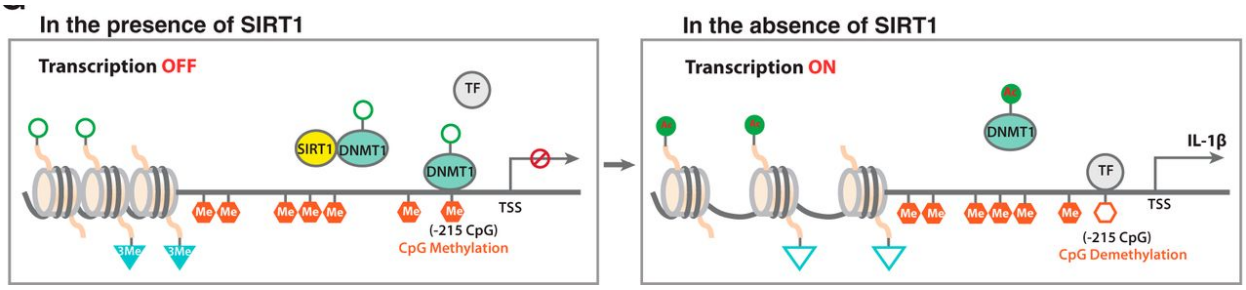


Figure 5-6: Potential mechanism of microglial SIRT1-mediated regulation of IL-1β transcript. With SIRT1, DNMT1 is deacetylated and activated. Activated DNMT1 methylates CpG at −215bp, recruits corepressor complex, and prevents the transcription factors (TFs) from binding to the promoter, and suppresses IL-1β transcription. Without SIRT1, DNMT1 remains acetylated and its activity is inhibited, leading to hypomethylation of IL-1β promoter at −215 bp, allowing TF binding.

others have used gene expression in peripheral blood to gain insights into the biology of neurodegenerative disorders (Coppola et al., 2011; Coppola et al., 2008). This study supports the notion that a disease-related signature is present in methylation data as well. Finally, we decided to focus on the risk-associated 17q21.31 region as an initial step, but many interesting candidates for further study emerged from the differential methylation analysis in PSP patients vs. controls, namely the nuclear transcription factor *NFYA*.

An additional follow-up study, driven by the hypothesis of SIRT1 deficiency contributing to aging by epigenetic regulation of specific genes, showed that SIRT1 levels in microglia exhibit an age-dependent decline, and microglial SIRT1 deficiency is causative in cognitive decline in normal aging and in FTD-related neurodegenerative diseases. Two probes, cg01290568 and cg15836722, were found to be significantly anti-correlated with aging in non-demented subjects. This pattern contradicts the majority of CpG sites in the genome, which become hypermethylated with age. Even more strikingly, these two probes were hypomethylated in dementia patients (with AD, FTD, FTD/MND, and PSP) compared to non-demented controls, even after correcting for age. Both of these findings are remarkably consistent with a model in which SIRT1 deficiency inhibits DNMT activity and reduces methylation of the proximal promoter of IL-1β, leading to IL-1β transcriptional upregulation (Figure 5-6). This finding

cements a central role for epigenetic regulation of gene expression in normal aging and neurodegenerative disease.

Although our analysis of published datasets supports the presence of the 17q21.31-associated methylation signature in brain tissues, further studies focused on brain samples from patients will be needed to test whether methylation changes are contributing to tissue-specific gene expression abnormalities, and ultimately explain the mechanism of action of genetic susceptibility alleles, and the striking regional vulnerability of these disorders.

Materials and Methods

Ethics statement. All subjects and/or their proxies signed informed consents for genetic studies. The research protocol was approved by the University of California San Francisco (UCSF) and Los Angeles (UCLA) University Institutional Review Boards for human research.

Sample description. Patients were enrolled as part of a large genetic study in neurodegenerative dementia (Genetic Investigation in Frontotemporal Dementia, GIFT) at the UCSF Memory and Aging Center (UCSF-MAC) (Coppola et al., 2007). 371 unrelated subjects were enrolled in the study (Supplementary Table C-1), including patients with neurodegenerative disorders (128 FTD, 43 PSP, and 15 AD), and 185 healthy controls.

Sample preparation. DNA was extracted from peripheral blood using standard methods. No cell sorting or cell selection was conducted, therefore our data measure methylation levels in whole blood. Total RNA was extracted from the same individuals from peripheral blood using Paxgene Blood RNA tubes (Qiagen).

Methylation arrays. Whole-genome methylation patterns were assayed by the Infinium Human Methylation450 BeadChip Kit (96 samples per chip). This work was performed in two stages (each including 2 chips), resulting in a total of 371 samples. Samples were hybridized as

follows: Dataset #1: PSP (n=40), FTD (n=55), Control (n=93); Dataset #2: FTD (n=73) AD (n=15), Control (n=92).

Genotyping. Taqman genotyping: Genotype variants at APOE (rs429358 and rs7412) and MAPT H1/H2 (rs1560310) were obtained using Taqman assays. Genome-wide SNP data was obtained using the Illumina HumanOmni1-Quad BeadChip. 17q21.31 sub-haplotypes were obtained by genotyping 6 SNPs as previously reported (Kauwe et al., 2008) using Taqman assays.

SNP arrays. High-throughput SNP genotyping data (Illumina HumanOmni1-Quad BeadChip) from a larger dataset containing 702 samples were available for 273 subjects (14 AD, 110 FTD, 15 PSP, and 134 Controls) in this study. SNP genotypes were called and exported from Illumina GenomeStudio (versions 1.6.3 and 1.8.4). Quality control included filtering for 1) SNPs with < 95% genotype call rates (n=157,121), 2) a minor allele frequency < 1% (n=126,502); and 3) with Hardy-Weinberg Equilibrium *p*-value < 1×10^{-6} in the control group (n=16,954). A total of 788,694 SNPs were included in the final analysis.

Ethnicity. We inferred ethnicity for 271 samples (out of the 273 for whom SNP data were available), by using SNP clustering compared to Hapmap data. Briefly, MDS analysis was applied on a merged dataset, including 702 samples from our data (273 of which were included in present study) and 1184 subjects from HapMap phase III. MDS plot shows that the first two principal components can cluster samples by ethnicity, and our data had good overlap with HapMap data (Figure S9). We used self-reported ethnicity for an additional 88 samples. For 12 samples ethnicity remained unknown either because we were not able to call ethnicity with certainty using SNP data (n=2), or because of the lack of SNP data and self-reported ethnicity (n=10).

Microarray-based gene expression analysis. Microarray expression data (Illumina HumanRef-8 v3.0) were available in 120 subjects with H1/H1 or H1/H2 haplotypes at 17q21.31.

Reduced representation bisulfite sequencing. RRBS was performed on seven samples multiplexed in two lanes of the Illumina HiSeq. Library preparation was performed using the Msp I restriction enzyme as previously described (Gu et al., 2011). Read alignment and methylation level calls were performed using BS-Seeker2 (Guo et al., 2013) (parameters: --aligner=bowtie -m 5 -g hg18.fa -r --low=50 --up=500 -a adapter.txt).

Published datasets. SNP data from Heyn *et al.* (2012) reporting the methylome analysis of newborns and centenarians, including 40 samples assayed by Illumina HumanMethylation450 BeadChip and 14 genotyped by Illumina HumanOmni5-Quad BeadChip (both methylation and SNP genotyping data were available for 12 samples), were downloaded from the Gene Expression Omnibus Database (GEO, http://www.ncbi.nlm.nih.gov/geo/, GSE31438). H1/H2 haplotypes were inferred from SNP rs1052553. For methylation data, beta and detection p-values were downloaded from GSE30870, and 65 sites containing missing values were removed.

Statistical Analysis

Methylation Arrays. In order to avoid potential confounders from batch effects, the two datasets were processed separately. Raw data was processed using the Illumina GenomeStudio software (version 2010.3). Background correction and color normalization were performed using the R package minfi version 1.2.0, and normalization using Subset-quantile Within Array Normalization (SWAN) (Maksimovic et al., 2012). Probes were excluded from further analysis if >95% samples had detection p-value >0.01. In summary, 3,027 probes were removed from dataset #1, and 26,306 probes were removed from dataset #2. In order to avoid potential confounders, 66,877 SNP-containing probes were also excluded from further analysis (Chen et al., 2013). Beta values (ratio between methylated probe intensity and the overall intensity) were computed using the R package minfi.

Linear models and empirical Bayes methods as implemented in the limma package were used for differential methylation analysis (Smyth, 2005). P-values were adjusted by using the Benjamini-Hochberg (BH) false discovery rate method. Multi-dimensional scaling (MDS) did not show obvious biases between chips within each dataset, but a batch effect could be observed between datasets #1 and #2 (Figure S10), similar to what has been reported in the literature (Bell et al., 2012; Zhang et al., 2010). To avoid potential confounders due to this batch effect, we compared each disease category with the set of controls within the same batch (i.e. conditioning on batch effect). Similar differential methylation results were obtained when we performed a combined analysis across the entire dataset, after correcting for batch effects using ComBat (Johnson et al., 2007) (Supplementary Material). Two filters were applied to conservatively identify differentially methylated sites: 1) a p-value-based filter (BH-adjusted $p \leq 0.05$) and 2) an absolute average beta difference (aβD) filter (absolute aβD >0.1). Chromosomal enrichment analysis was performed by using the hypergeometric test as implemented in the R phyper function.

The impact of relative cell counts in peripheral blood was estimated as previously described (Feinberg, 2010; Portela and Esteller, 2010) and based on a subset (n=385) of the 500 loci whose methylation levels reflect the relative proportions of immune cells in unfractionated whole blood. After estimating the blood cell type distribution for each sample using the methylation level of the 385 loci, we applied a linear mixed-effect model considering (1) main blood cell types distribution as dependent variables; (2) disease status (or 17q21.31 haplotype), age, ethnicity, and gender as fixed effects; and (3) chip number as a random effect (Supplementary Material).

Raw and normalized methylation data were deposited in the Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo), accession number: GSE53740.

Methylation QTL analysis (methQTL) was performed by regressing methylation level at select CpG sites of interest on SNP genotypes. Age and the first two principal component generated

from MDS analysis were also included in the multivariate regression model as covariates. Linear regression methQTL analysis was performed using PLINK (Purcell et al., 2007).

Microarray-based expression analysis. Raw data were processed as previously described (Coppola, 2011). Briefly, after quantile normalization, batch effects were removed using ComBat (Johnson et al., 2007). Differential expression analysis was performed using the limma package (Smyth, 2005), applying a false discovery rate filter of ≤ 0.05 and an absolute log fold change filter of >0.1.

Causality analysis was performed using the software package Network Edge Orienting (NEO) (Aten et al., 2008), a structural equation modeling software for determining the direction of causality among various phenotypes (e.g. clinical, molecular) given genotype data. Subjects from all batches were included (n = 219). The relative fitting index of the model is estimated by the Single Marker LEO.NB score, defined as the base-10 logarithm of the probability ratio between the mediation model and the next most likely causal model (i.e., a LEO.NB score of 1 means that the fit of the mediation model is 10 times better than that of the next best alternative causal model). The genotype was encoded as the dosage of the minor allele (A) at rs1560310 tagging H2 (i.e., 0 for H1/H1, 1 for H1/H2, and 2 for H2/H2). Gene expression levels were included for differentially expressed probes according to the 17q21.31 haplotype, and were encoded with the ComBat-corrected relative expression levels. Clinical phenotype was encoded as a binary variable (i.e., 0 for unaffected, 1 for affected with PSP). The Single-Marker analysis option was used, and results surpassing the thresholds of a LEO.NB score (a likelihood ratio of model fit) > 0.8 and RMSEA index < 0.05 were considered significant fits to the mediation model (Horvath, 2011).

References

Akbarian, S., Beeri, M., and Haroutunian, V. (2013). Epigenetic determinants of healthy and diseased brain aging and cognition. JAMA Neurology *70*, 711-718.

Aten, J.E., Fuller, T.F., Lusis, A.J., and Horvath, S. (2008). Using genetic markers to orient the edges in quantitative trait networks: the NEO software. BMC Systems Biology *2*, 34.

Baker, M., Litvan, I., Houlden, H., Adamson, J., Dickson, D., Perez-Tur, J., Hardy, J., Lynch, T., Bigio, E., and Hutton, M. (1999). Association of an Extended Haplotype in the Tau Gene with Progressive Supranuclear Palsy. Human Molecular Genetics *8*, 711-715.

Bell, J.T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A.*, et al.* (2012). Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. PLoS Genetics *8*, e1002629.

Boeve, B.F. (2012). Progressive supranuclear palsy. Parkinsonism & Related Disorders *18, Supplement 1*, S192-S194.

Caffrey, T.M., Joachim, C., Paracchini, S., Esiri, M.M., and Wade-Martins, R. (2006). Haplotype-specific expression of exon 10 at the human MAPT locus. Human Molecular Genetics *15*, 3529-3537.

Caffrey, T.M., Joachim, C., and Wade-Martins, R. (2008). Haplotype-specific expression of the N-terminal exons 2 and 3 at the human MAPT locus. Neurobiology of Aging *29*, 1923-1929.

Caffrey, Tara M., and Wade-Martins, R. (2012). The role of MAPT sequence variation in mechanisms of disease susceptibility. Biochemical Society Transactions *40*, 687-692.

Chen, Y.-a., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., and Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics *8*, 203-209.

Cho, S.-H., Chen, J.A., Sayed, F., Ward, M.E., Gao, F., Nguyen, T.A., Krabbe, G., Sohn, P.D., Lo, I., Minami, S.*, et al.* (2015). SIRT1 Deficiency in Microglia Contributes to Cognitive Decline in Aging and Neurodegeneration via Epigenetic Regulation of IL-1β. The Journal of Neuroscience *35*, 807-818.

Conrad, C., Andreadis, A., Trojanowski, J.Q., Dickson, D.W., Kang, D., Chen, X., Wiederholt, W., Hansen, L., Masliah, E., Thal, L.J.*, et al.* (1997). Genetic evidence for the involvement of ? in progressive supranuclear palsy. Annals of Neurology *41*, 277-281.

Coppola, G. (2011). Designing, Performing, and Interpreting a Microarray-Based Gene Expression Study. In Neurodegeneration: Methods and Protocols, G. Manfredi, and H. Kawamata, eds. (Totowa, NJ: Humana Press), pp. 417-439.

Coppola, G., Burnett, R., Perlman, S., Versano, R., Gao, F., Plasterer, H., Rai, M., Saccá, F., Filla, A., Lynch, D.R.*, et al.* (2011). A gene expression phenotype in lymphocytes from friedreich ataxia patients. Annals of Neurology *70*, 790-804.

Coppola, G., Chinnathambi, S., Lee, J.J., Dombroski, B.A., Baker, M.C., Soto-Ortolaza, A.I., Lee, S.E., Klein, E., Huang, A.Y., Sears, R.*, et al.* (2012). Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. Human Molecular Genetics *21*, 3500-3512.

Coppola, G., Karydas, A., Rademakers, R., Wang, Q., Baker, M., Hutton, M., Miller, B.L., and Geschwind, D.H. (2008). Gene expression study on peripheral blood identifies progranulin mutations. Annals of Neurology *64*, 92-96.

Coppola, G., Miller, B.L., Chui, H., Varpetian, A., Levey, A., Cotman, C.W., DeCarli, C., Mendez, M.F., Bartzokis, G., and Kukull, W.A. (2007). Genetic investigation in frontotemporal dementia and Alzheimer's disease: the GIFT Study. Annals of Neurology *62*, S52.

Davies, M.N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., Harris, R.A., Milosavljevic, A., Troakes, C.*, et al.* (2012). Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biology *13*, R43.

Evans, W., Fung, H.C., Steele, J., Eerola, J., Tienari, P., Pittman, A., Silva, R.d., Myers, A., Vrieze, F.W.-D., Singleton, A.*, et al.* (2004). The tau H2 haplotype is almost exclusively Caucasian in origin. Neuroscience Letters *369*, 183-185.

Feinberg, A.P. (2010). Epigenomics reveals a functional genome anatomy and a new approach to common disease. Nat Biotech *28*, 1049-1052.

Gan, L., and Mucke, L. (2008). Paths of Convergence: Sirtuins in Aging and Neurodegeneration. Neuron *58*, 10-14.

Ghosh, S., Yates, A.J., Frühwald, M.C., Miecznikowski, J.C., Plass, C., and Smiraglia, D. (2010). Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. Epigenetics *5*, 527-538.

Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.-L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J.*, et al.* (2010). Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain. PLoS Genetics *6*, e1000952.

Groemping, U. (2006). Relative Importance for Linear Regression in R: The Package relaimpo. Journal of Statistical Software *17*, 27.

Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protocols *6*, 468-481.

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.-Y., and Pellegrini, M. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics *14*, 774.

Hayesmoore, J.B.G., Bray, N.J., Cross, W.C., Owen, M.J., O'Donovan, M.C., and Morris, H.R. (2009). The effect of age and the H1c MAPT haplotype on MAPT expression in human brain. Neurobiology of Aging *30*, 1652-1656.

Hernandez, D.G., Nalls, M.A., Gibbs, J.R., Arepalli, S., van der Brug, M., Chong, S., Moore, M., Longo, D.L., Cookson, M.R., Traynor, B.J.*, et al.* (2011). Distinct DNA methylation changes highly correlated with chronological age in the human brain. Human Molecular Genetics *20*, 1164-1172.

Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J.V., Setien, F., Carmona, F.J.*, et al.* (2012). Distinct DNA methylomes of newborns and centenarians. Proceedings of the National Academy of Sciences *109*, 10522-10527.

Hoglinger, G.U., Melhem, N.M., Dickson, D.W., Sleiman, P.M.A., Wang, L.-S., Klei, L., Rademakers, R., de Silva, R., Litvan, I., Riley, D.E.*, et al.* (2011). Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nature Genetics *43*, 699-705.

Horvath, S. (2011). Weighted Network Analysis (New York: Springer).

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biology *14*, 3156.

Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86.

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A.*, et al.* (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature *393*.

Jakovcevski, M., and Akbarian, S. (2012). Epigenetic mechanisms in neurological disease. Nature Medicine *18*, 1194-1204.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.

Kalinderi, K., Fidani, L., and Bostantjopoulou, S. (2009). From 1997 to 2007: A decade journey through the H1 haplotype on 17q21 chromosome. Parkinsonism & Related Disorders *15*, 2-5.

Kauwe, J.S.K., Cruchaga, C., Mayo, K., Fenoglio, C., Bertelsen, S., Nowotny, P., Galimberti, D., Scarpini, E., Morris, J.C., Fagan, A.M.*, et al.* (2008). Variation in MAPT is associated with cerebrospinal fluid tau levels in the presence of amyloid-beta deposition. Proceedings of the National Academy of Sciences of the United States of America *105*, 8050-8054.

Koestler, D.C., Christensen, B.C., Karagas, M.R., Marsit, C.J., Langevin, S.M., Kelsey, K.T., Wiencke, J.K., and Houseman, E.A. (2013). Blood-based profiles of DNA methylation predict the underlying distribution of cell types. Epigenetics *8*, 816-826.

Ladd-Acosta, C., Pevsner, J., Sabunciyan, S., Yolken, R.H., Webster, M.J., Dinkins, T., Callinan, P.A., Fan, J.-B., Potash, J.B., and Feinberg, A.P. (2007). DNA Methylation Signatures within the Human Brain. The American Journal of Human Genetics *81*, 1304-1315.

Libert, S., and Guarente, L. (2013). Metabolic and neuropsychiatric effects of calorie restriction and sirtuins. Annual Review of Physiology *75*, 669-684.

Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M.*, et al.* (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotech *31*, 142-147.

Lu, H., Liu, X., Deng, Y., and Qing, H. (2013). DNA methylation, a hand behind neurodegenerative diseases. Frontiers in Aging Neuroscience *5*.

Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome Biology *13*, R44.

Michan, S., and Sinclair, D. (2007). Sirtuins in mammals: insights into their biological function. Biochemical Journal *404*, 1-13.

Myers, A.J., Pittman, A.M., Zhao, A.S., Rohrer, K., Kaleem, M., Marlowe, L., Lees, A., Leung, D., McKeith, I.G., Perry, R.H.*, et al.* (2007). The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. Neurobiology of Disease *25*, 561-570.

Pearl, J. (2009). Causality, 2nd Edition (Cambridge: Cambridge University Press).

Pittman, A.M., Myers, A.J., Abou-Sleiman, P., Fung, H.C., Kaleem, M., Marlowe, L., Duckworth, J., Leung, D., Williams, D., Kilford, L.*, et al.* (2005). Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. Journal of Medical Genetics *42*, 837-846.

Pittman, A.M., Myers, A.J., Duckworth, J., Bryden, L., Hanson, M., Abou-Sleiman, P., Wood, N.W., Hardy, J., Lees, A., and de Silva, R. (2004). The structure of the tau haplotype in controls and in progressive supranuclear palsy. Human Molecular Genetics *13*, 1267-1274.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat Biotech *28*, 1057-1068.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J.*, et al.* (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. American Journal of Human Genetics *81*, 559-575.

Rademakers, R., Melquist, S., Cruts, M., Theuns, J., Del-Favero, J., Poorkaj, P., Baker, M., Sleegers, K., Crook, R., De Pooter, T.*, et al.* (2005). High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy. Human Molecular Genetics *14*, 3281-3292.

Schadt, E.E., Lamb, J., Yang, X., Zhu, J., Edwards, S., GuhaThakurta, D., Sieberts, S.K., Monks, S., Reitman, M., Zhang, C.*, et al.* (2005). An integrative genomics approach to infer causal associations between gene expression and disease. Nature Genetics *37*, 710-717.

Simon-Sanchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G.*, et al.* (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nature Genetics *41*, 1308-1312.

Smyth, G.K. (2005). Limma: linear models for microarray data. In Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, eds. (New York: Springer), pp. 397–420.

Steele, J.C., Richardson, J., and Olszewski, J. (1964). Progressive supranuclear palsy: A heterogeneous degeneration involving the brain stem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. Archives of Neurology *10*, 333-359.

Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V.G.*, et al.* (2005). A common inversion under selection in Europeans. Nature Genetics *37*, 129-137.

Trabzuni, D., Wray, S., Vandrovcova, J., Ramasamy, A., Walker, R., Smith, C., Luk, C., Gibbs, J.R., Dillman, A., Hernandez, D.G.*, et al.* (2012). MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. Human Molecular Genetics *21*, 4094-4103.

Urdinguio, R.G., Sanchez-Mut, J.V., and Esteller, M. (2009). Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. The Lancet Neurology *8*, 1056-1072.

van Eijk, K.R., de Jong, S., Boks, M.P., Langeveld, T., Colas, F., Veldink, J.H., de Kovel, C.G., Janson, E., Strengman, E., Langfelder, P.*, et al.* (2012). Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. BMC Genomics *13*, 636.

Vansteelandt, S., and Lange, C. (2012). Causation and causal inference for genetic effects. Human Genetics *131*, 1665-1676.

Vilariño-Güell, C., Soto-Ortolaza, A.I., Rajput, A., Mash, D.C., Papapetropoulos, S., Pahwa, R., Lyons, K.E., Uitti, R.J., Wszolek, Z.K., Dickson, D.W.*, et al.* (2011). MAPT H1 haplotype is a risk factor for essential tremor and multiple system atrophy. Neurology *76*, 670-672.

Wade-Martins, R. (2012). Genetics: The MAPT locus[mdash]a genetic paradigm in disease susceptibility. Nat Rev Neurol *8*, 477-478.

Williams, D.R., and Lees, A.J. (2009). Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges. The Lancet Neurology *8*, 270-279.

Zhang, D., Cheng, L., Badner, J.A., Chen, C., Chen, Q., Luo, W., Craig, D.W., Redman, M., Gershon, E.S., and Liu, C. (2010). Genetic Control of Individual Differences in Gene-Specific Methylation in Human Brain. The American Journal of Human Genetics *86*, 411-419.

Zhang, F., Wang, S., Gan, L., Vosler, P.S., Gao, Y., Zigmond, M.J., and Chen, J. (2011). Protective effects and mechanisms of sirtuins in the nervous system. Progress in Neurobiology *95*, 373-395.

Zhu, J., Wiener, M.C., Zhang, C., Fridman, A., Minch, E., Lum, P.Y., Sachs, J.R., and Schadt, E.E. (2007). Increasing the Power to Detect Causal Associations by Combining Genotypic and Expression Data in Segregating Populations. PLoS Computational Biology *3*, e69.

Zoghbi, H.Y. (2009). Rett syndrome: what do we know for sure? Nature Neuroscience *12*, 239-240.

Chapter 6: Genome-wide association study of plasma tau concentrations

Introduction

Tau is a microtubule-associated protein (MAP) that promotes assembly and stabilization of cytoskeletal microtubules (Lee et al., 2001). Accumulation of insoluble deposits of tau has been linked to the pathogenesis of a number of neurodegenerative disorders, including Alzheimer's disease (AD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), chronic traumatic encephalopathy, and some variants of frontotemporal lobar degeneration, which are collectively classified as tauopathies (Lee et al., 2001).

Cerebrospinal fluid (CSF) tau concentrations are thought to reflect neuronal degeneration in AD, and CSF tau, either alone or in combination with beta-Amyloid peptide ($A\beta_{42}$), has been confirmed as a useful biomarker for AD across the spectrum of disease severity (Jack and Holtzman, 2013). Based on this, CSF total tau (t-tau) and phosphorylated tau at residue 181 (p-tau) concentrations have been used as endophenotypes in genome-wide association studies (GWAS) to detect risk variants for AD, with the *APOE* locus showing the strongest association with elevated CSF tau (Cruchaga et al., 2013; Kim et al., 2011). Enigmatically, despite strong genetic links to tau, CSF tau levels are normal or low in other tauopathies such as PSP and CBD, and some tau gene (*MAPT*) mutation carriers (Wagshal et al., 2015). The recent development of an ultra-sensitive assay for tau in peripheral blood makes it feasible to study the relationship between peripheral tau concentrations and tauopathies (Olivera et al., 2015). By comparison to CSF, plasma tau levels in aging and neurodegenerative disease have not been well studied.

The use of quantitative traits in GWAS has been shown to increase statistical power over case-control designs (Cruchaga et al., 2013; Kim et al., 2011). Here, we hypothesized that plasma tau, similar to CSF tau, may constitute a suitable endophenotype for a GWAS designed to identify genetic factors involved in tau metabolism and highlighting relevant tau-related

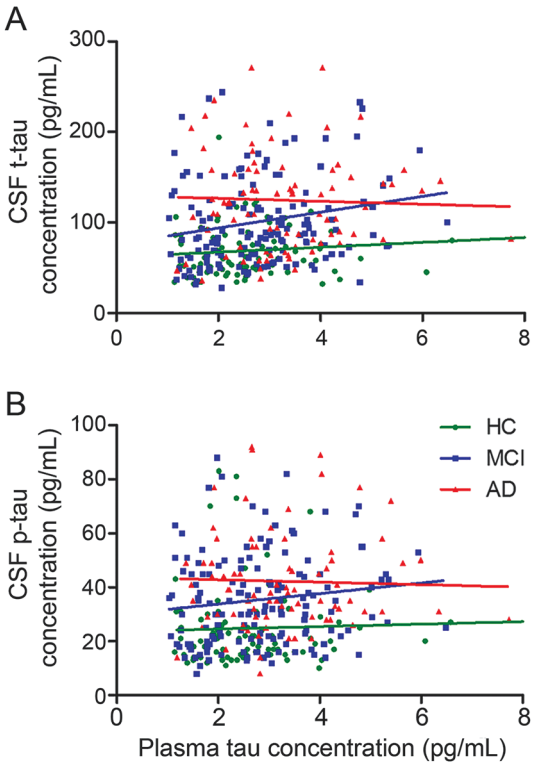


Figure 6-1: Correlations between plasma tau and CSF tau levels. (a) There was no correlation between plasma tau levels and CSF total tau in any diagnostic group. **(b)** There was no correlation between plasma tau levels and CSF phosphorylated tau in any diagnostic group. AD = Alzheimer disease; HC = healthy controls; MCI = mild cognitive impairment.

physiological and pathophysiological processes. Within this context, we conducted a GWAS for plasma tau level and identified a single nucleotide polymorphism (SNP) (rs242557) within the *MAPT* gene that showed a genome-wide significant association with plasma--but not CSF--tau levels.

Correlations between plasma tau and CSF tau levels

There were 316 (AD = 83, MCI = 149, HC = 84) subjects with both plasma and CSF tau levels. However, there was no correlation between tau levels in plasma and CSF in any diagnostic group [plasma tau versus CSF t-tau (Figure 6-1a): AD (r = 0.131, *P* = 0.234), MCI (r = 0.209, *P* = 0.066), HC (r = 0.002, *P* = 0.984); plasma tau versus CSF p-tau (Figure 6-1b): AD (r = 0.056, *P* = 0.611), MCI (r = 0.145, *P* = 0.077), HC (r = -0.027, *P* = 0.809)], indicating that plasma tau levels do not reflect CSF tau levels.

Table 6-1: Demographic Information

Baseline diagnosis	AD	MCI	HC	Total
n	149	163	151	463
Age, mean ± SD (range), y	76.1 ± 7.3 (56.5-90.9)	74.6 ± 7.6 (54.4-88.8)	75.7 ± 4.9 (62.0-89.6)	75.5 ± 6.7 (54.4-90.9)
M/F, n	83/66	110/53	89/62	282/181
APOE ε4 carrier, %	67.1	53.4	25.8	49.0
Plasma tau, mean ± SD, pg/mL[a]	3.2 ± 1.3	2.9 ± 1.3	2.7 ± 1.1	2.9 ± 1.2

Abbreviations: AD = Alzheimer disease; HC = healthy controls; MCI = mild cognitive impairment.
[a] Plasma tau concentrations were different across the 3 diagnostic groups ($p = 0.006$). Post hoc pairwise analyses after Bonferroni correction showed that only patients with AD had higher plasma tau levels compared with HC ($p = 0.004$).

MAPT but not APOE genotype is associated with plasma tau levels

463 individuals (AD = 149, MCI = 163, HC = 151, at baseline) were identified for GWAS (Table 6-1). Plasma tau concentrations were different between the three diagnostic groups (p = 0.006) (Table 6-1). Post hoc analysis after Bonferroni correction showed that only AD patients had higher plasma tau levels compared to controls (P = 0.004). After adjusting for age, gender and diagnosis, a genome-wide significant association of rs242557 (in the *MAPT* region) with elevated plasma tau levels ($P = 1.59 \times 10^{-8}$) was detected (Table 6-2, Figure 6-2a). This locus survived both permutation-based and Bonferroni corrections for multiple testing (empirical P [EMP1]= 0.0002; permutation-based corrected empirical P [EMP2] = 0.005; Bonferroni corrected P = 0.005). The minor risk allele (A) of rs242557 was associated with higher plasma tau levels in a dose-dependent effect within both combined and each diagnostic group (normal group, $P = 5.9\times 10^{-4}$; MCI group, P =0.002; and AD group, P =0.001) (Figure 6-3a).

There were no other genome-wide significant associations with plasma tau outside the *MAPT* region. In the *MAPT* region, several SNPs in linkage disequilibrium (LD) with rs242557 showed p-values lower than 0.01 for plasma tau levels (Figure 6-2b). However, after controlling for rs242557 genotype (Figure 6-2c), no SNPs in this region showed association with plasma tau levels, indicating that all the association in this locus was driven by rs242557.

Table 6-2: Top SNPs associated with plasma tau

CHR	SNP	Observed MAF	Closest gene	SNP type/location	p, Plasma tau	Empiric FWER
17	rs242557	0.372	*MAPT*	Intron	4.85×10^{-9}	0.0024
6	rs2187213	0.359	*PARK2*	Intron	6.15×10^{-6}	0.7562
10	rs7072793	0.393	*IL2RA*	Upstream +1978bp	7.89×10^{-6}	0.8334
10	rs7073236	0.394	*IL2RA*	Upstream +2264bp	7.89×10^{-6}	0.8334
9	rs7047280	0.404	*ELAVL2*	Intergenic	8.13×10^{-6}	0.8394

Abbreviations: bp = base pair; CHR = chromosome; FWER = family-wise error rate; MAF = minor allele frequency; SNP = single nucleotide polymorphism. rs7072793 is in complete linkage disequilibrium with rs7073236 ($r^2 = 1$).

The LD pattern between rs242557 and nearby SNPs was nearly identical in the ADNI cohort compared with 1000 Genomes European subjects (Supplementary Figure D-2) suggesting that the SNP genotypes from this study were accurate. Moreover, LD between rs242557 and an H1/H2 haplotype-defining SNP (rs1560310) was calculated using the 1000 Genomes Project EUR cohort (r^2 = 0.197 and D' = 1), demonstrating that rs242557 is specific to the H1 clade.

The analysis identified two other suggestive loci, *IL2RA* (rs7072793 and rs7073236) and *PARK2* (rs2187213), where p-values reached the level of $P < 10^{-6}$ (Table 6-2, Figure 6-2a). Both *IL2RA* and *PARK2* minor alleles were associated with lower plasma tau levels in a dose-dependent effect within both combined and each diagnostic group (Figure 6-3b and Figure 6-3c). However, they did not survive after both permutation-based and Bonferroni corrections for multiple testing. Other plasma tau associated SNPs that did not reach genome-wide significance and whose *P* values are between 10^{-6} and 10^{-5} are listed in Supplementary Table D-1.

APOE but not MAPT locus affects CSF, but not plasma tau levels

Among the 463 individuals analyzed in the plasma tau levels, there were 314 (AD = 82, MCI = 148, HC = 84) subjects with CSF tau levels. We investigated whether top SNPs identified in the plasma tau concentration GWAS (rs242557 within *MAPT*) and previous CSF tau concentration GWAS (*APOE*) associated with CSF tau levels. After adjusting for age, gender and diagnosis,

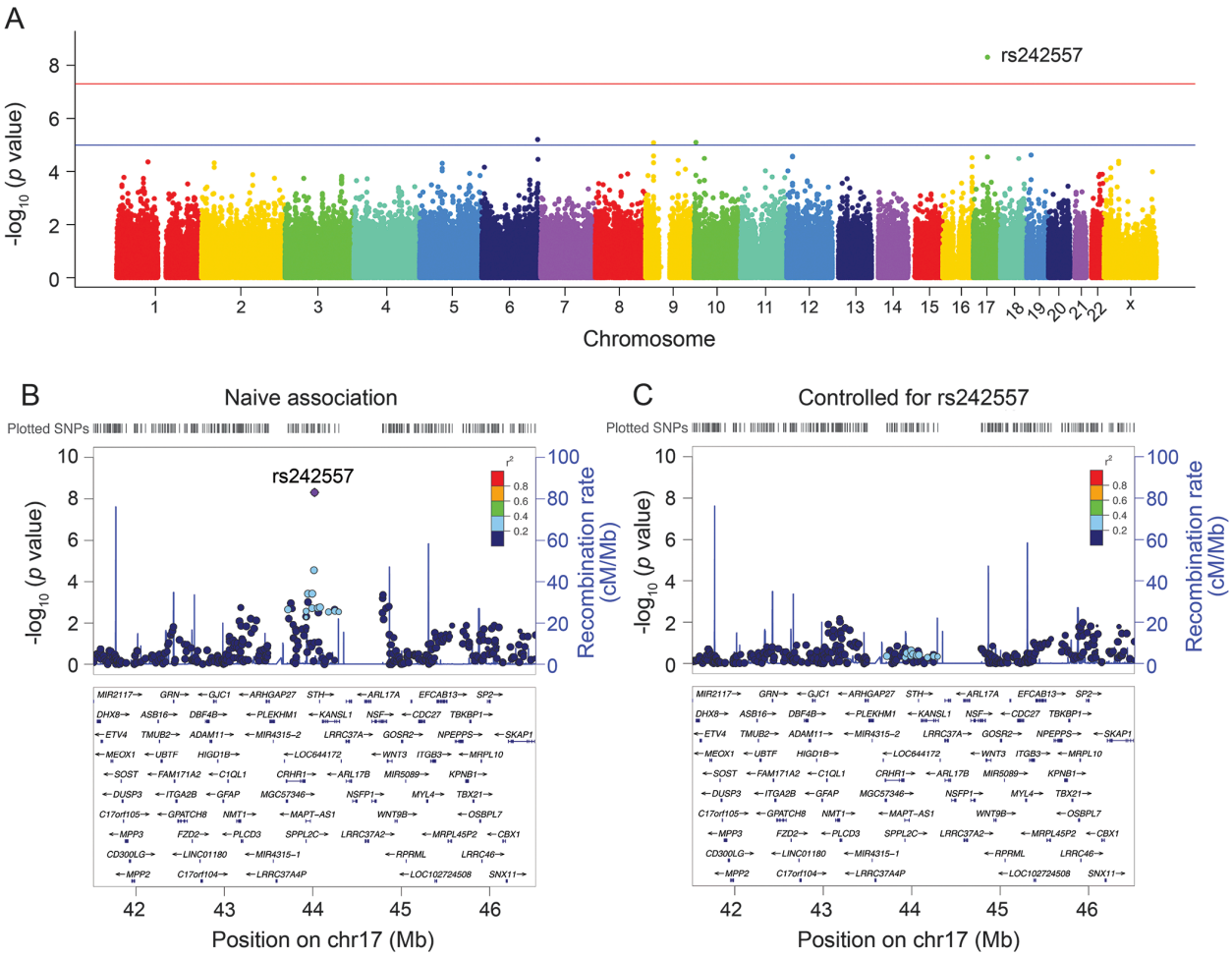


Figure 6-2: Manhattan and regional plots for associations with plasma tau. (a) Genome-wide signal intensity (Manhattan) plots showing the -log$_{10}$ (*p* value) for individual single nucleotide polymorphisms. **b)** Regional association results for the *MAPT* region of chromosome 17. **(c)** Association results for 17q21.31 controlling for rs242557.

APOE ε4 showed significant associations with both CSF t-tau (*P* = 0.004) and CSF p-tau (*P* = 0.028) after Bonferroni corrections. However, rs242557 within *MAPT* was not associated with CSF t-tau (*P* = 0.988) or CSF p-tau (*P* = 0.835). *APOE* ε4 was not associated with plasma tau levels (*P* = 0.312), nor were other SNPs in the *APOE* region.

Discussion

To our knowledge, the present study constitutes the first GWAS of plasma tau levels in a large sample. We identified a genome-wide significant association of a SNP in the tau gene (*MAPT)* region with plasma tau levels and two additional suggestive association loci (in *IL2RA* and *PARK2*). The minor allele (A) of *MAPT* H1c (rs242557) was associated with higher plasma tau

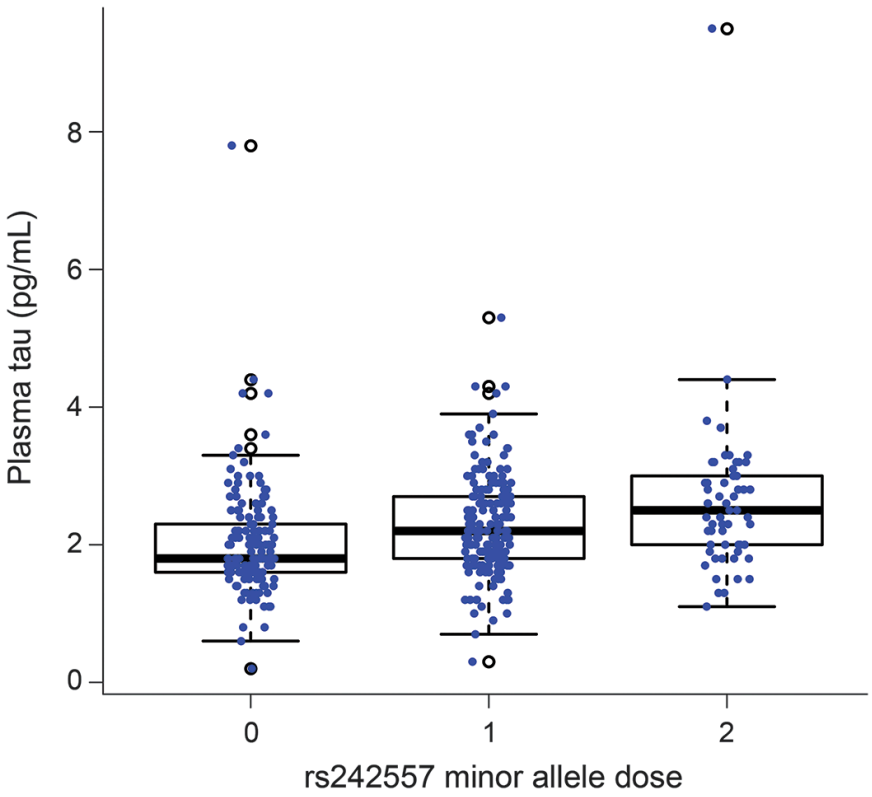


Figure 6-3: Plasma tau levels in a replication cohort as a function of rs242557 genotype. Plasma tau levels were compared across the GG, GA, and AA genotypes of rs242557 in an independent cohort of 387 participants to validate the initially observed association. A significant association of increasing plasma tau concentration with increasing minor allele (A) dose of rs242557 was observed ($p = 1.0 \times 10^{-5}$).

levels in a dose-dependent fashion, whereas both *IL2RA* (rs7072793 and rs7073236) and *PARK2* (rs2187213) minor alleles were associated with lower plasma tau levels.

Genetic alterations in and around *MAPT* are strong risk factors for neurodegenerative tauopathies. The *MAPT* gene locus is located on chromosome 17q21 (Pittman et al., 2005; Pittman et al., 2004). It exists as two major haplotype groups, termed H1 and H2, with the majority of individuals having the H1/H1 haplotype. Up to 25% of individuals in Western populations have a ~ 970 Kb sequence, including *MAPT*, oriented in the reverse orientation, inducing a larger 1.3-1.6 Mb region of linkage disequilibrium (LD) (Pittman et al., 2004; Zody et al., 2008). Genetic studies, including GWAS, have identified both the inversion polymorphism and haplotype-specific polymorphisms influencing the risk of 4-repeat (4R) tauopathies (PSP

and CBD) (Baker et al., 1999; Conrad et al., 1997; Kouri et al., 2015; Pittman et al., 2005; Pittman et al., 2004; Rademakers et al., 2005; Williams et al., 2007). Moreover, the common sub-haplotype H1c (tagged by rs242557) on the background of the H1 has been consistently found to be associated with both PSP and CBD (Hoglinger et al., 2011; Kouri et al., 2015; Pittman et al., 2005; Williams et al., 2007). Previously published GWAS data with neuropathologically diagnosed cases showed that rs242557, representing the *MAPT* H1c haplotype, was one of the most common SNPs associated with PSP (odds ratio/OR = 1.96, $P = 4.2 \times 10^{-70}$) and CBD (OR = 1.57, $P = 7.91 \times 10^{-6}$) (Table 6-2) (Hoglinger et al., 2011; Kouri et al., 2015). However, rs242557 was not associated with the risk of AD ($P = 0.974$) (Allen et al., 2014), which is pathologically characterized by intracellular neurofibrillary tangles composed of equal ratios of 3R and 4R tau (Lee et al., 2001).

The mechanism by which MAPT H1c haplotype could increase plasma tau levels is not clear. The SNP rs242557 falls into a 182-bp highly conserved region that was previously identified as a key regulatory polymorphism influencing *MAPT* expression (Figure 6-2b) (Caffrey et al., 2006; Majounie et al., 2013; Myers et al., 2007; Pittman et al., 2005; Rademakers et al., 2005; Trabzuni et al., 2012). In cultured cells, the rs242557 A allele with the H1 background promoter variant had 2.7-fold greater transcriptional activity than allele G ($P < 0.01$) with the H1 background promoter variant and 4.2-fold greater than allele G ($P < 0.01$) with the H2 background promoter variant (Myers et al., 2007). It has also been hypothesized that rs242557 could affect *MAPT* splicing (Caffrey et al., 2006; Hoglinger et al., 2011). The *MAPT* H1c risk haplotype has been reported to be associated with increased mRNA expression of tau and 4 repeat containing transcripts (Majounie et al., 2013; Myers et al., 2007). Consistent with these data, this risk allele has also been associated with higher *MAPT* expression levels in the cerebellum and temporal cortex from AD autopsy subjects (Allen et al., 2014). Our study included data from normal controls and patients with AD spectrum disorders, but interestingly

found a strong association with a genetic risk factor previously implicated in non-AD tauopathies (CBD and PSP). Therefore, we cannot rule out an interaction between AD pathology and the H1c haplotype in our sample that may partially account for the association with plasma tau. However the subgroup analysis which showed that rs242557 was associated with plasma tau levels in a dose-dependent effect within each diagnostic group (normal controls, MCI and AD; Figure 6-2), suggests that the presence of AD pathology is not necessary to observe the plasma tau association.

Elevated plasma tau was not explained by elevated CSF tau. No correlations between tau levels in plasma and CSF were identified (Figure 6-1) suggesting that different mechanisms are likely to regulate plasma tau concentrations. Previous GWAS have determined that the *APOE* locus is the strongest association for CSF tau and ptau levels (Cruchaga et al., 2013; Kim et al., 2011). Although rs242557 was associated with higher CSF tau protein levels in one small sample of AD patients (n =89) (Laws et al., 2007), both our data (n=314) and another study with 313 individuals did not identified any association of this SNP with CSF tau protein levels (Kauwe et al., 2008). Multiple *MAPT* loci have been strongly associated with PSP and CBD, and some autosomal dominant *MAPT* mutations, particularly those in IVS10, are known to produce a PSP-like syndrome (Morris et al., 2003). Together these data indicate that plasma tau levels might be a more useful endophenotype for identifying genetic risk for 4-repeat tauopathies (PSP and CBD), than for AD.

The suggestive alleles that were identified in the plasma tau GWAS might also be associated with risk for tauopathies. Mutations in the *PARK2* gene, which encodes the parkin RBR E3 ubiquitin protein ligase, are the most common cause of early-onset parkinsonism (Lucking et al., 2000; Schrag and Schott, 2006). Interestingly, a *PARK2* polymorphism (Val380Leu) is associated with lower risk of PSP while another mutation (Cys212Tyr) produces clinical and

pathological features of PSP (Morales et al., 2002; Ros et al., 2008). *IL2RA,* which encodes the interleukin 2 (IL2) receptor alpha, is a multiple sclerosis susceptibility gene and plays an important role in regulating immune response (International Multiple Sclerosis Genetics et al., 2007). Moreover, an *IL2* SNP (rs6852535) (P = 1.3× 10^{-7}) was identified as a suggestive locus for PSP risk in a previous GWAS study (Hoglinger et al., 2011). Recent data also suggest that microglia may play a role in tau related neurodegeneration, which would be consistent with this association between an immunological risk factor gene and plasma tau (Asai et al., 2015). Further studies will be required to identify the potential roles of *PARK2* and *IL2RA* in tauopathies.

A limitation of this report is the modest sample size for a GWAS, which precluded stratified analyses for each diagnostic group. We applied a more stringent MAF threshold (MAF > 0.20) and a stringent correction of individual *P* values (both permutation-based and Bonferroni corrections), which may have excluded less common SNPs associated with plasma tau levels. Moreover, the minor allele of rs242557 showed a dose-dependent effect specifically with plasma tau levels, which further indicates that the detected SNP within *MAPT* gene represents biological signal and not analysis artifacts or type 1 error.

In summary, we detected a genome-wide significant SNP, rs242557 in *MAPT*, and two suggestive loci (in *IL2RA* and *PARK2*) associated with plasma tau levels measured in healthy elders and individuals with MCI or AD. Our study results suggest that plasma tau is a valid endophenotype in GWAS and can be used to characterize the metabolism and functions of tau. As rs242557 represents the *MAPT* H1c haplotype that has previously been identified as a major genetic risk factor for both PSP and CBD, our findings also suggest that plasma tau concentration could be a useful endophenotype for identifying risk for 4-repeat tauopathies. Replication studies with independent, larger datasets will be required to confirm these findings.

Methods

Subjects. In this study, 463 (AD = 149, mild cognitive impairment, MCI = 163, healthy controls, HC = 151, at baseline) non-Hispanic Caucasian individuals whose data met all quality control (QC) criteria were included from the Alzheimer's Disease Neuroimaging Initiative 1 (ADNI 1) cohort. Table 6-1 shows the demographic data and description of the plasma tau levels in each group. Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

The full cohort with both plasma tau and GWAS data included 506 subjects. The analysis was restricted to non-Hispanic Caucasian participants (n = 468) to reduce the potential bias of population stratification in the GWAS. Cryptic relatedness and population substructure, which can confound GWAS, were checked with genomic identity-by-descent (IBD) and multidimensional scaling (MDS) components (Supplementary Figure D-1). This step removed two participants who seemed to be cryptically related and clustering separately from the other samples (Supplementary Figure D-1a), resulting in 466 valid samples. Finally, using the HapMap cohort, they showed tight clustering with individuals of European ancestry (Supplementary Figure D-1b).

Plasma and CSF tau measurements and quality control. Plasma tau concentrations were determined using the Human Total Tau kit from Quanterix (Quanterix, Lexington, MA, USA), as described by the manufacturer. Assays were run at the University of Gothenburg. Intra- and

inter-assay coefficients of variation were 10-15%. The lower limit of quantification was 1.22 ng/L. The plasma tau concentrations could be downloaded from the ADNI1 database (http://adni.loni.usc.edu/). Detailed steps for measurements and QC using CSF t-tau and CSF p-tau have been previously reported (Kim et al., 2011). Further QC was performed to reduce the potential influence of extreme outliers on statistical results. Mean and standard deviations (SD) baseline plasma tau measures were calculated, blind to diagnostic information and subjects who had a value greater or smaller than 4 SD from the mean value (7.7ng/L) were regarded as extreme outliers and removed from the analysis (Kim et al., 2011). This step removed three additional participants, resulting in 463 valid samples.

Standard protocol approvals, registrations, and patient consents. The study was approved by institutional review boards of all participating institutions and written informed consent was obtained from all participants or authorized representatives.

Genotyping and quality control. The ADNI 1 samples were genotyped with the Human 610-Quad BeadChip (Illumina, Inc., San Diego, CA). Stringent QC assessment was performed using the PLINK software with the following criteria: minimum call rate for SNPs and individuals >98%, minimum minor allele frequencies (MAF) > 0.20, Hardy-Weinberg equilibrium test $P > 0.001$. The restriction to SNPs with a MAF greater than 20% served to reduce the likelihood of false-positive results to enhance statistical power. rs7412 and rs429358, which define the apolipoprotein E (*APOE*) alleles, were genotyped separately by an *APOE* genotyping kit (Kim et al., 2011).

Statistical analyses. Spearman rank correlation coefficient was used to determine correlations of plasma tau concentrations with CSF tau concentrations. $P < 0.05$ was considered statistically significant after adjustment for multiple comparisons using Bonferroni correction. Association of plasma tau with the genetic variants was performed using PLINK with the additive genetic

model, i.e., dose-dependent effect of the minor allele. IBD and MDS were calculated in PLINK. The analysis included a total of 305,283 genotyped variants. Age, gender and diagnosis were included as covariates. We considered an association as genome-wide significant at a threshold of $P = 5x10^{-8}$. To further exclude possible false-positive results, PLINK's max (T) permutation test with 5,000 permutations was used to generate empirical p-values and for multiple testing correction. The effects of genotypes on plasma tau levels were also examined with a multiple linear regression model.

References

Allen, M., Kachadoorian, M., Quicksall, Z., Zou, F., Chai, H.S., Younkin, C., Crook, J.E., Pankratz, V.S., Carrasquillo, M.M., Krishnan, S.*, et al.* (2014). Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels. Alzheimers Res Ther *6*, 39.

Asai, H., Ikezu, S., Tsunoda, S., Medalla, M., Luebke, J., Haydar, T., Wolozin, B., Butovsky, O., Kugler, S., and Ikezu, T. (2015). Depletion of microglia and inhibition of exosome synthesis halt tau propagation. Nat Neurosci.

Baker, M., Litvan, I., Houlden, H., Adamson, J., Dickson, D., Perez-Tur, J., Hardy, J., Lynch, T., Bigio, E., and Hutton, M. (1999). Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum Mol Genet *8*, 711-715.

Caffrey, T.M., Joachim, C., Paracchini, S., Esiri, M.M., and Wade-Martins, R. (2006). Haplotype-specific expression of exon 10 at the human MAPT locus. Hum Mol Genet *15*, 3529-3537.

Conrad, C., Andreadis, A., Trojanowski, J.Q., Dickson, D.W., Kang, D., Chen, X., Wiederholt, W., Hansen, L., Masliah, E., Thal, L.J.*, et al.* (1997). Genetic evidence for the involvement of tau in progressive supranuclear palsy. Ann Neurol *41*, 277-281.

Cruchaga, C., Kauwe, J.S., Harari, O., Jin, S.C., Cai, Y., Karch, C.M., Benitez, B.A., Jeng, A.T., Skorupa, T., Carrell, D.*, et al.* (2013). GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. Neuron *78*, 256-268.

Hoglinger, G.U., Melhem, N.M., Dickson, D.W., Sleiman, P.M., Wang, L.S., Klei, L., Rademakers, R., de Silva, R., Litvan, I., Riley, D.E.*, et al.* (2011). Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat Genet *43*, 699-705.

International Multiple Sclerosis Genetics, C., Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B.*, et al.* (2007). Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med *357*, 851-862.

Jack, C.R., Jr., and Holtzman, D.M. (2013). Biomarker modeling of Alzheimer's disease. Neuron *80*, 1347-1358.

Kauwe, J.S., Cruchaga, C., Mayo, K., Fenoglio, C., Bertelsen, S., Nowotny, P., Galimberti, D., Scarpini, E., Morris, J.C., Fagan, A.M.*, et al.* (2008). Variation in MAPT is associated with cerebrospinal fluid tau levels in the presence of amyloid-beta deposition. Proc Natl Acad Sci U S A *105*, 8050-8054.

Kim, S., Swaminathan, S., Shen, L., Risacher, S.L., Nho, K., Foroud, T., Shaw, L.M., Trojanowski, J.Q., Potkin, S.G., Huentelman, M.J., *et al.* (2011). Genome-wide association study of CSF biomarkers Abeta1-42, t-tau, and p-tau181p in the ADNI cohort. Neurology *76*, 69-79.

Kouri, N., Ross, O.A., Dombroski, B., Younkin, C.S., Serie, D.J., Soto-Ortolaza, A., Baker, M., Finch, N.C., Yoon, H., Kim, J., *et al.* (2015). Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. Nat Commun *6*, 7247.

Laws, S.M., Friedrich, P., Diehl-Schmid, J., Muller, J., Eisele, T., Bauml, J., Forstl, H., Kurz, A., and Riemenschneider, M. (2007). Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease. Mol Psychiatry *12*, 510-517.

Lee, V.M., Goedert, M., and Trojanowski, J.Q. (2001). Neurodegenerative tauopathies. Annu Rev Neurosci *24*, 1121-1159.

Lucking, C.B., Durr, A., Bonifati, V., Vaughan, J., De Michele, G., Gasser, T., Harhangi, B.S., Meco, G., Denefle, P., Wood, N.W., *et al.* (2000). Association between early-onset Parkinson's disease and mutations in the parkin gene. N Engl J Med *342*, 1560-1567.

Majounie, E., Cross, W., Newsway, V., Dillman, A., Vandrovcova, J., Morris, C.M., Nalls, M.A., Ferrucci, L., Owen, M.J., O'Donovan, M.C., *et al.* (2013). Variation in tau isoform expression in different brain regions and disease states. Neurobiol Aging *34*, 1922 e1927-1922 e1912.

Morales, B., Martinez, A., Gonzalo, I., Vidal, L., Ros, R., Gomez-Tortosa, E., Rabano, A., Ampuero, I., Sanchez, M., Hoenicka, J., *et al.* (2002). Steele-Richardson-Olszewski syndrome in a patient with a single C212Y mutation in the parkin protein. Mov Disord *17*, 1374-1380.

Morris, H.R., Osaki, Y., Holton, J., Lees, A.J., Wood, N.W., Revesz, T., and Quinn, N. (2003). Tau exon 10 +16 mutation FTDP-17 presenting clinically as sporadic young onset PSP. Neurology *61*, 102-104.

Myers, A.J., Pittman, A.M., Zhao, A.S., Rohrer, K., Kaleem, M., Marlowe, L., Lees, A., Leung, D., McKeith, I.G., Perry, R.H., *et al.* (2007). The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. Neurobiol Dis *25*, 561-570.

Olivera, A., Lejbman, N., Jeromin, A., French, L.M., Kim, H.S., Cashion, A., Mysliwiec, V., Diaz-Arrastia, R., and Gill, J. (2015). Peripheral Total Tau in Military Personnel Who Sustain Traumatic Brain Injuries During Deployment. JAMA Neurol.

Pittman, A.M., Myers, A.J., Abou-Sleiman, P., Fung, H.C., Kaleem, M., Marlowe, L., Duckworth, J., Leung, D., Williams, D., Kilford, L.*, et al.* (2005). Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. J Med Genet *42*, 837-846.

Pittman, A.M., Myers, A.J., Duckworth, J., Bryden, L., Hanson, M., Abou-Sleiman, P., Wood, N.W., Hardy, J., Lees, A., and de Silva, R. (2004). The structure of the tau haplotype in controls and in progressive supranuclear palsy. Hum Mol Genet *13*, 1267-1274.

Rademakers, R., Melquist, S., Cruts, M., Theuns, J., Del-Favero, J., Poorkaj, P., Baker, M., Sleegers, K., Crook, R., De Pooter, T.*, et al.* (2005). High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy. Hum Mol Genet *14*, 3281-3292.

Ros, R., Ampuero, I., and Garcia de Yebenes, J. (2008). Parkin polymorphisms in progressive supranuclear palsy. J Neurol Sci *268*, 176-178.

Schrag, A., and Schott, J.M. (2006). Epidemiological, clinical, and genetic characteristics of early-onset parkinsonism. Lancet Neurol *5*, 355-363.

Trabzuni, D., Wray, S., Vandrovcova, J., Ramasamy, A., Walker, R., Smith, C., Luk, C., Gibbs, J.R., Dillman, A., Hernandez, D.G.*, et al.* (2012). MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. Hum Mol Genet *21*, 4094-4103.

Wagshal, D., Sankaranarayanan, S., Guss, V., Hall, T., Berisha, F., Lobach, I., Karydas, A., Voltarelli, L., Scherling, C., Heuer, H.*, et al.* (2015). Divergent CSF tau alterations in two common tauopathies: Alzheimer's disease and progressive supranuclear palsy. J Neurol Neurosurg Psychiatry *86*, 244-250.

Williams, D.R., Pittman, A.M., Revesz, T., Lees, A.J., and de Silva, R. (2007). Genetic variation at the tau locus and clinical syndromes associated with progressive supranuclear palsy. Mov Disord *22*, 895-897.

Zody, M.C., Jiang, Z., Fung, H.C., Antonacci, F., Hillier, L.W., Cardone, M.F., Graves, T.A., Kidd, J.M., Cheng, Z., Abouelleil, A.*, et al.* (2008). Evolutionary toggling of the MAPT 17q21.31 inversion region. Nat Genet *40*, 1076-1083.

Chapter 7: Tauopathies and the Future of Drug Discovery

Science is the captain and practice the soldiers.

- Leonardo da Vinci, *Notebooks* – edited by Richter & Wells (da Vinci, 2008)

The inchoate crisis of neurodegenerative disease burden (particularly Alzheimer's disease, AD) looms large over society; the thought of losing one's memories, mobility, and life strikes fear into the most stoic individuals. Despite intense research efforts, progress in therapeutics has not met the burgeoning need. Few genes (outside of tau) have been implicated in tauopathies, and theories regarding the disease mechanisms are still incomplete and controversial. Even the effects of the H1 haplotype of chr17q21.31, a well-established risk factor for progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD), are still poorly characterized. My research attempted to clarify the genetics and thus the pathobiology of the tauopathies, with the ultimate objective to enable development of efficacious treatments.

Discovery of new genes that cause tauopathy

Genetic risk factors can take many forms: common variants that confer slight risk for disease; structural variants that cannot be typed by traditional means; rare variants that require large sample sizes to observe effects; variants that have not yet been found; and everything in between. Discovery of novel genetic risk factors in each category requires a specialized statistical approach. In a series of studies, I systematically dissected the genetics of tauopathies with particular attention to maximizing coverage of the types of variants interrogated.

To cover common variants (typically of small effect), I performed a genome-wide association study. I established a genetic overlap between PSP, Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS), suggesting that some of the pathogenic mechanisms of tauopathies are shared between neurodegenerative diseases. Performing a joint analysis to combine our

genome-wide association studies with existing data, I identified novel associations with loci near the genes *RUNX2*, *SLCO1A2*, *DUSP10*, *SP1*, *ASAP1*, and *WDR63*. Further work was performed to fine-map and to predict which genes in these loci were true causal genes. From these arrays, I was also able to call copy number variation, and identified that copy number variants of tau may cause PSP.

More "uncommon" variants, in particular those with facile interpretation in the exome, could be accessed with the newly-developed exome array. The exome array provides a middle ground between the coverage of rare variants in genome sequencing and the scalability and low cost of traditional array genotyping. Because of limited power for variant-wise analysis of the exome array. I employed and developed statistical methods to collapse variants for analysis at the gene level. Among known disease genes, I identified that functional exonic variants in *ABCA7* may contribute to AD, a finding that has since been confirmed in multiple studies (Del-Aguila et al., 2015; Le Guennec et al., 2016; Steinberg et al., 2015). I also identified the novel disease genes for AD, *PAXIP1* and *DYSF*.

The ultimate look at the genome comes from genome sequencing. While we have not identified novel disease genes, likely due to lack of statistical power, we have found hints of a signal and pinpoint the likely locations of disease-causing mutations. Furthermore, our sample cohort provided definitive confirmation of the role of a rare coding variant of tau in PSP, with one of the largest observed enrichments for the *MAPT* A152T allele observed to date. Work to call copy number variation is still ongoing, and a larger sample cohort with more statistical power is currently being assembled.

As more of the disease mechanisms become known and as sample cohorts (and corresponding statistical power) increase, new avenues appear for discovering genetic risk factors. Future

work will leverage disease mechanisms, improved phenotypic and neuropathologic stratification of patients, and knowledge about the function of genetic variants to further increase statistical power to find new causal genes. We lay the groundwork here for a truly integrative framework to continue to expand the boundaries of knowledge.

Elucidating the mechanisms of mutations in tau

The role of even the few identified genetic risk factors in disease is unclear. In PSP, the mystery of the H1 haplotype in particular captivated my interest. This polymorphism is highly prevalent in the population and has been toggled multiple times through evolution, suggesting perhaps that it is selected for. And yet, it is somehow overrepresented in patients with PSP. The haplotype overlaps the gene encoding tau, yet is not associated with pathogenic protein coding mutations or drastic shifts in expression level (there are, however, slight changes in splicing, which we and others have identified for exon 3).

To explain the risk conferred, we have identified a methylation signature correlated to the chromosome 17q21.31 risk haplotype. Applying the mathematics of causal inference using structural equation models, the most parsimonious explanation was that the methylation status (at some of the methylated sites) mediated the risk for the allele. This epigenetic marker, more so than even the measured expression levels, may provide clues as to the true role of the risk haplotype in disease. This role for methylation was not unique to the chromosome 17q21.31 locus. In collaboration with Li Gan at the Gladstone Institutes, I looked at promoter methylation of the gene encoding IL-1β (*IL1B)* in this cohort. Her group had identified this regulatory marker as a downstream target of sirtuin 1 in mouse models. In agreement with those findings, I found in humans that the methylation (at sites homologous to those identified in mouse) decreased with age (and even more so with tauopathies). The extent of promoter methylation also appeared to control mRNA expression of *IL1B*. Methylation is an important regulatory marker

that may mediate drastic genetic influences on the course of disease, and future work in brain-relevant cell types and additional epigenetic markers may even further illuminate the mechanisms of disease.

Forging a path for drug discovery and precision medicine

Translating the insights of genomics to drug discovery has flummoxed many on both sides of the science-industry divide. And it is a divide, each side with drastically divergent epistemology, traditions, and ultimate objectives. I believe in the inevitability that genomics will disrupt how we treat diseases and develop therapeutics, and even how we define diseases. As a foray into achieving these goals and potentially bridging that divide, I started Verge Genomics with a good friend and colleague in the Medical Scientist Training Program, Alice Zhang. We founded the company to identify the genomic changes correlated with disease, and discover ways to normalize them as treatment – guided by science and serving the patients. While straightforward in ideology, the challenge (and where many have failed) is to recognize that non-disease-causal factors drive the majority of variation in maps of gene expression, epigenetic markers, etc. To see behind these murky waters has proven highly non-trivial.

In "The Innovator's Prescription", Christensen and colleagues describe a three-stage succession of the state of medical treatment (Christensen et al., 2008). The first stage subsumes treatments based upon a physician's *intuition*. As medical science advances, the field progresses into a second stage characterized by probabilistic guidelines based on *empirical* data, such as randomized controlled trials. Eventually, in the third stage, the causes of disease and methods for diagnosis of particular subgroups of each disease are established; as efficacious therapies can be targeted to disease mechanisms with *precision*, treatment outcomes can be streamlined and optimized. As the mechanisms of disease crystallize, physicians can apply the powerful methods of deduction to problems that previously could only

yield to induction. The current clinical management of tauopathies largely falls in the first two stages; progress hinges upon improvements in understanding the precise disease mechanisms and methods to classify patients. This work furthers the advance into the third stage in three ways: 1) discovery of new genes that cause tauopathies, laying the groundwork for identifying distinct biology behind similar clinical presentations; 2) elucidating the downstream pathogenic effects of genetic variants, particularly the chromosome 17q21.31 haplotype containing tau; and 3) development of methods to bring to bear on drug discovery, some of which are already being applied to practice.

To illustrate some of these points in greater depth, I relay the case of a patient of Dr. Adam Boxer that I was fortunate to observe at the University of California, San Francisco. A charming woman approaching the eighth decade of life reported frequent falls and weakness in her hands. On exam she was alert and oriented, her cognitive faculties completely intact. Her posture was somewhat awkward and her countenance had become gloomy, but she could not hide an irrepressible zeal for life. I noticed a striking vertical gaze palsy, rigidity, and left-sided ideomotor apraxia. Her gait was unsteady, and seemed unable to shift weight to her right foot. We felt that she had signs and history characteristic of CBD or PSP, which she had suspected as well. She cried at the thought of this potentially terminal diagnosis. The tenor of the situation changed, however, when Dr. Boxer adeptly noticed small patches of white matter hyperintensities on FLAIR, in a somewhat periventricular distribution. He postulated that multiple sclerosis (for which treatments are available) could be responsible for her symptoms, and additional studies confirmed the diagnosis. Without knowledge of disease mechanisms gained from the MRI, we would be unable to treat the patient correctly or even understand what disease processes are at work. While perhaps not as divergent as MS, the individual tauopathies are undoubtedly composed of even finer divisions of disease pathobiology. For example, even within FTD, the neuropathology often does not even include tau. In the future,

genomics screens based on findings like the ones presented here may be analogous to the transformative role of the MRI to precisely diagnose and treat diseases, and I am grateful to have had the opportunity to make contributions to this exciting area.

I have been extraordinarily fortunate to work with Profs. Coppola, Geschwind, Lazareff, and Horvath in the vanguard of scientific discovery and medical research. While this dissertation punctuates a milestone in my education, it is merely a starting point for the field, and we march inexorably onward to improve the lives of our patients now and in the future.

References

Christensen, C., Grossman, J.H., and Hwang, J. (2008). The Innovator's Prescription (New York: McGraw-Hill).

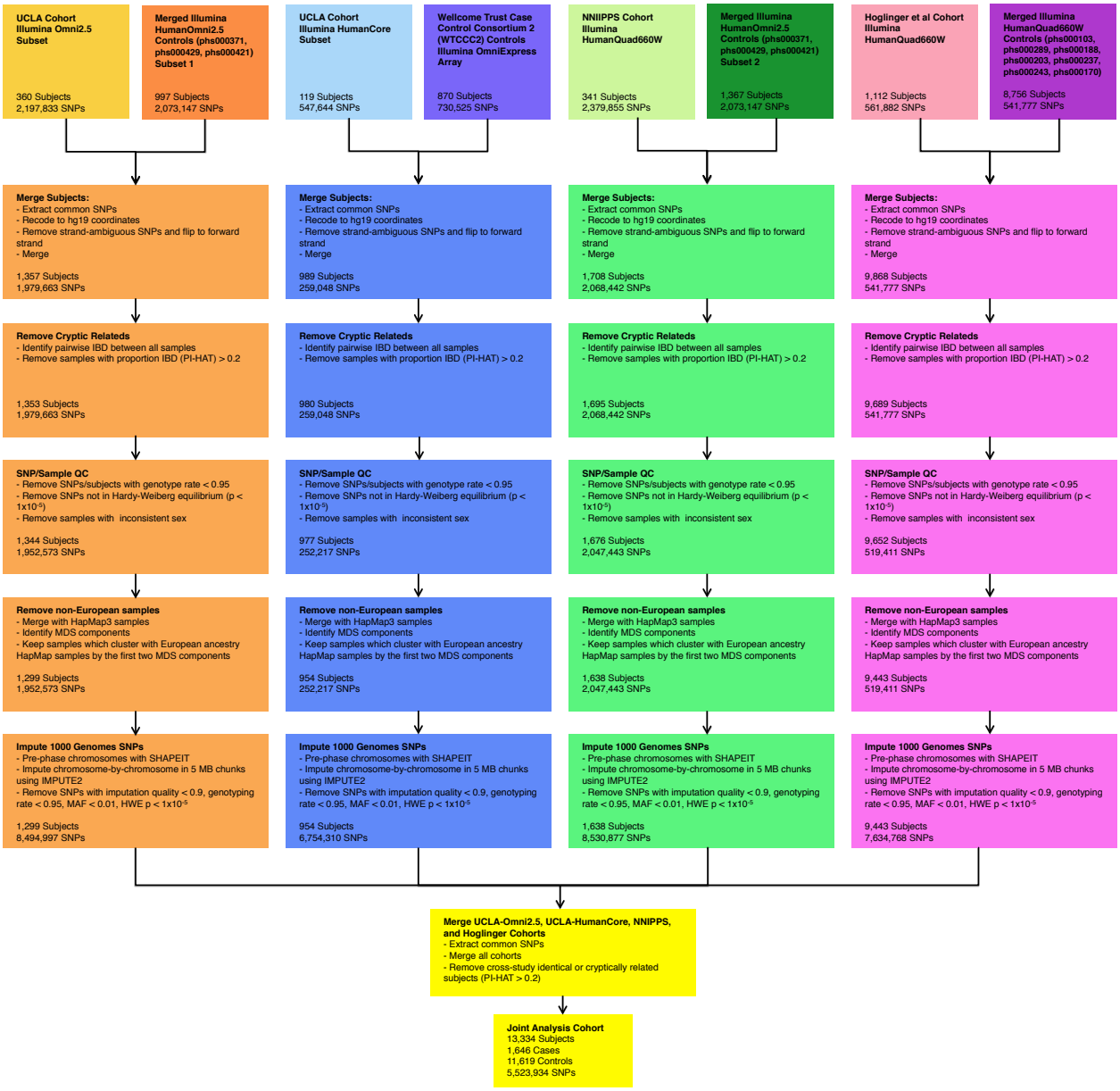
da Vinci, L. (2008). Notebooks (Oxford: Oxford University Press).

Del-Aguila, J.L., Fernández, M.V., Jimenez, J., Black, K., Ma, S., Deming, Y., Carrell, D., Saef, B., Howells, B., Budde, J.*, et al.* (2015). Role of ABCA7 loss-of-function variant in Alzheimer's disease: a replication study in European–Americans. Alzheimer's Research and Therapy *7*, 73.

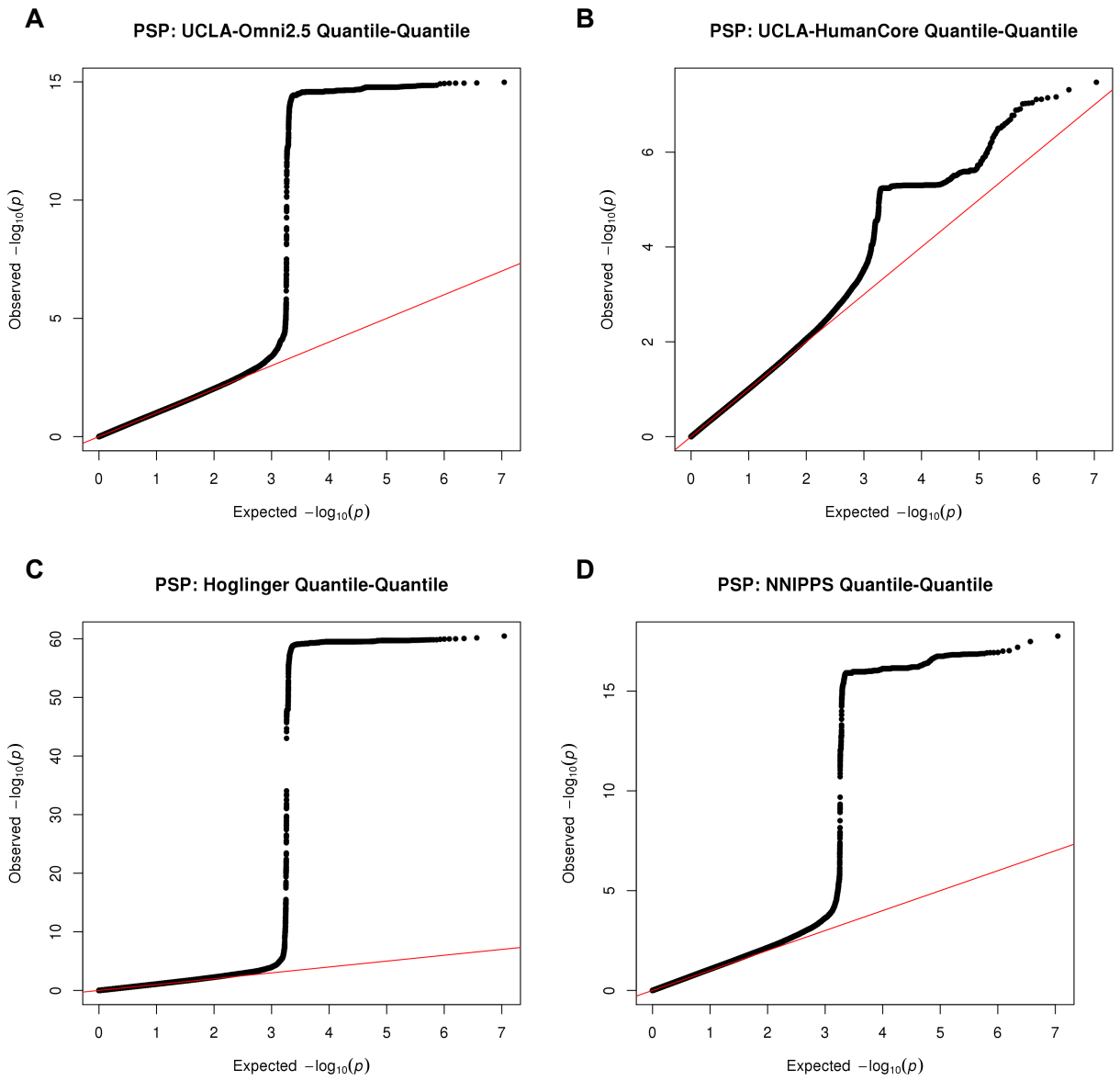
Le Guennec, K., Nicolas, G., Quenez, O., Charbonnier, C., Wallon, D., Bellenguez, C., Grenier-Boley, B., Rousseau, S., Richard, A.-C., Rovelet-Lecrux, A.*, et al.* (2016). ABCA7 rare variants and Alzheimer disease risk. Neurology *86*, 2134-2137.

Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H., Sulem, P., Magnusson, O.T., Gudjonsson, S.A., Unnsteinsdottir, U.*, et al.* (2015). Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nature Genetics *47*, 445-447.

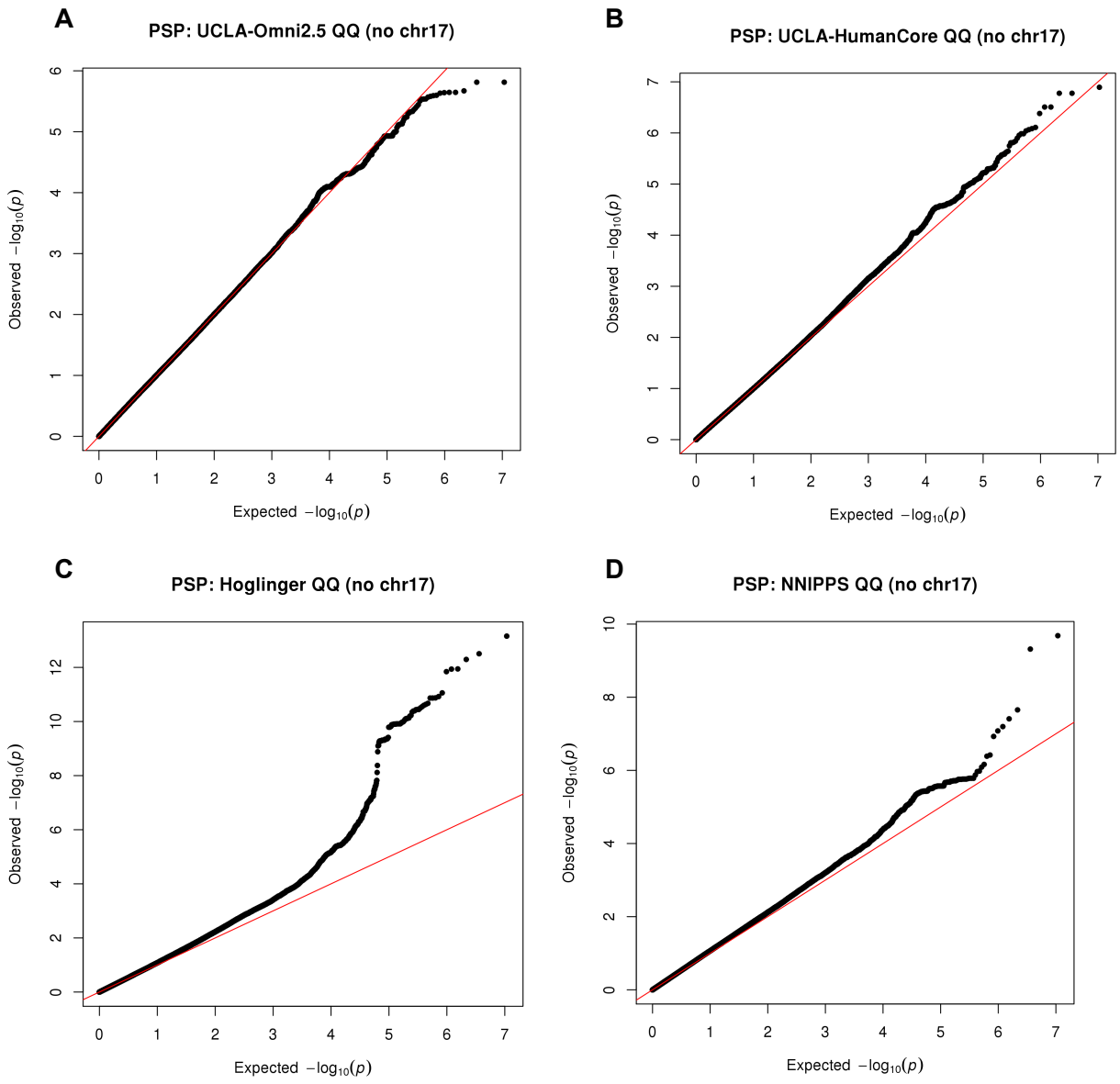
Appendix A: Supplementary Material for Chapter 2



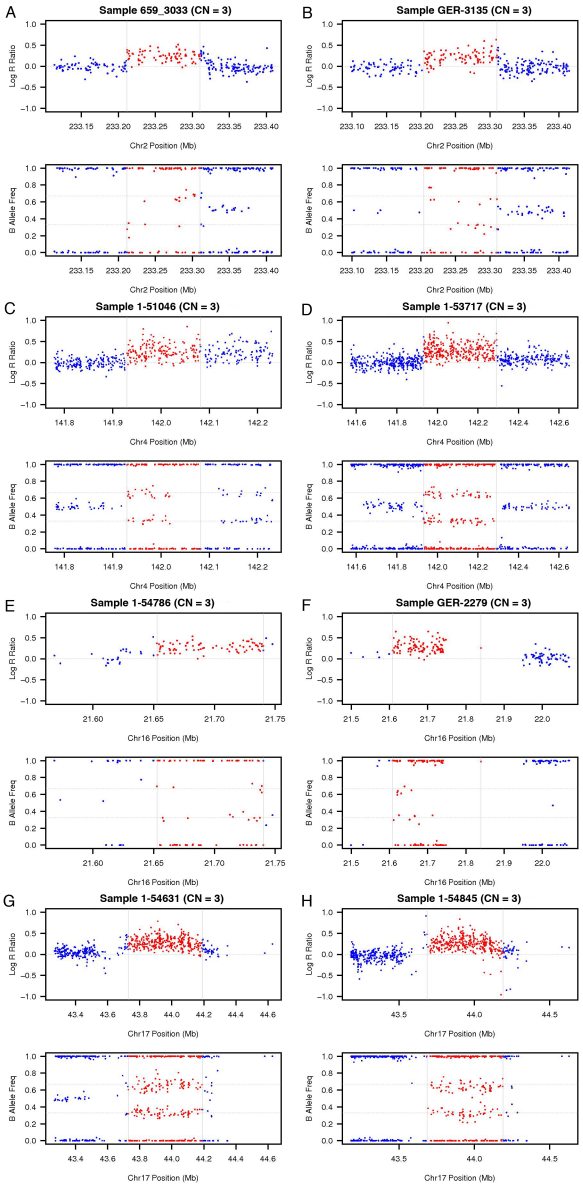
Supplementary Figure A-1: Pre-processing steps to create each individual study cohort and the merged joint analysis cohort.



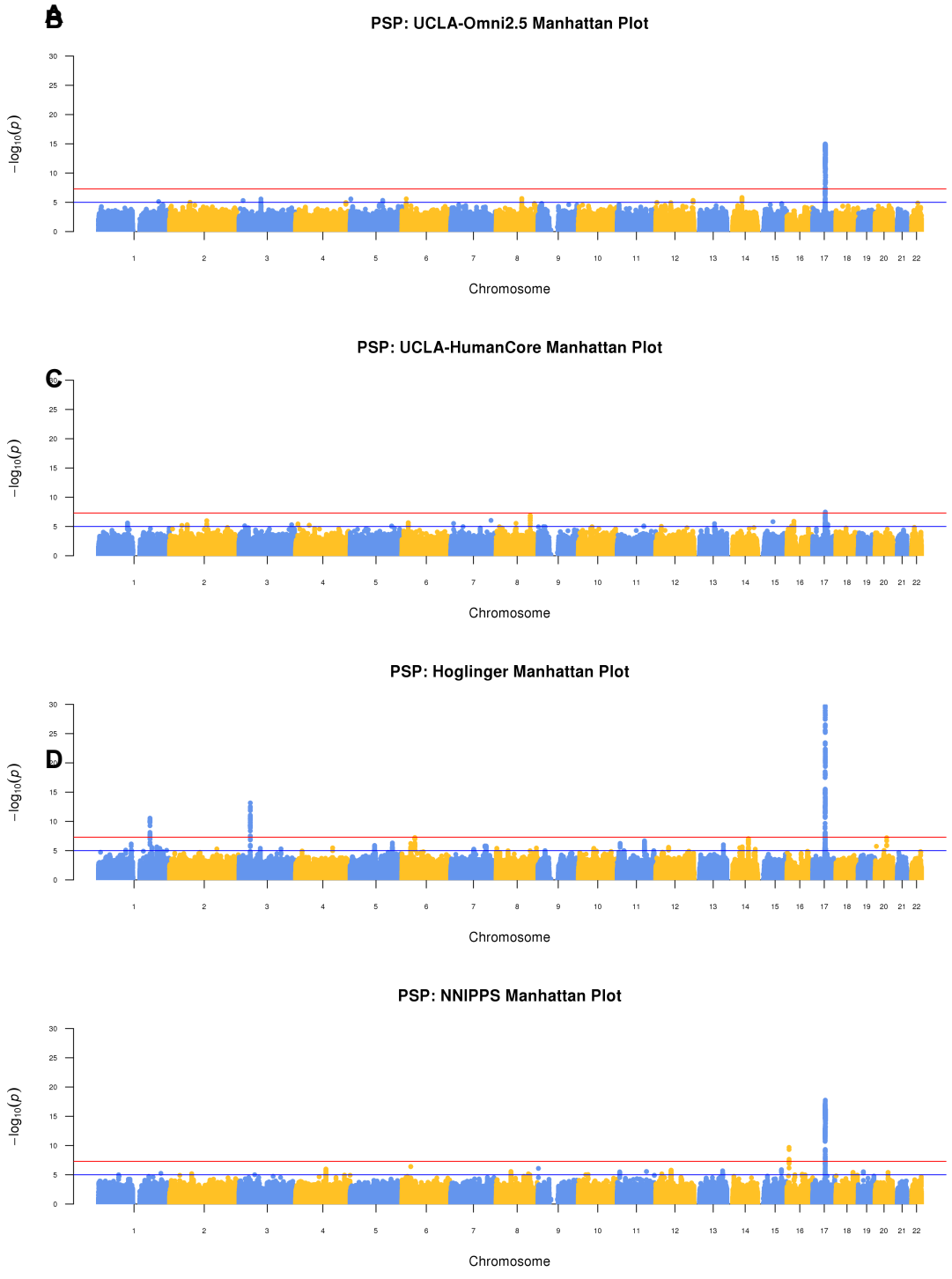
Supplementary Figure A-2: Quantile-Quantile (QQ) plots showing the potential for test statistic inflation in A) the UCLA Omni 2.5 cohort; B) the UCLA HumanCore cohort; C) the Hoglinger cohort; and D) the NNIPPS cohort.



Supplementary Figure A-3: Quantile-Quantile (QQ) plots, excluding chromosome 17, in A) the UCLA Omni 2.5 cohort; B) the UCLA HumanCore cohort; C) the Hoglinger cohort; and D) the NNIPPS cohort. Because of the strong association with an extended haplotype region in 17q21, a disproportionate number of associated SNPs reside on chromosome 17.

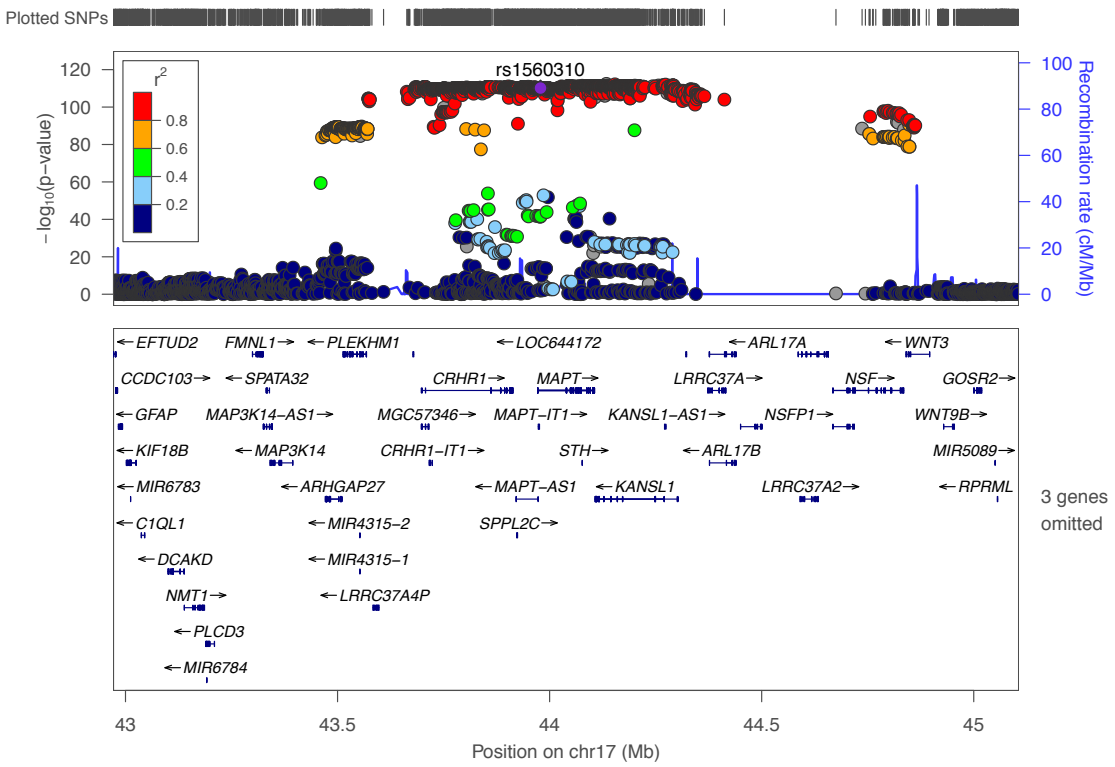


Supplementary Figure A-4: Signal intensity of recurrent rare CNV calls in PSP cases. Log R Ratio and B Allele Frequency are shown for CNVs in 2q37.1 (A and B), 4q31.21 (C and D), 16p12.2 (E and F), and 17q21.31 (G and H).

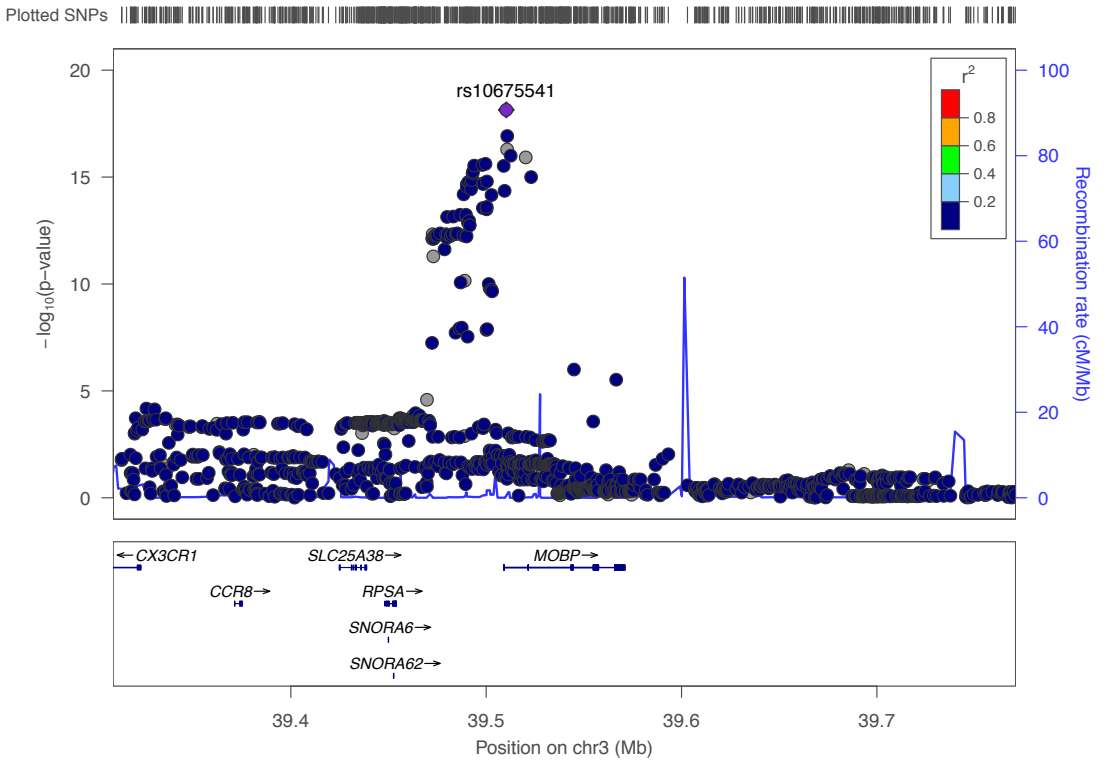


Supplementary Figure A-5: Genome-wide association results by genomic position in A) the UCLA Omni 2.5 cohort; B) the UCLA HumanCore cohort; C) the Hoglinger cohort; and D) the NNIPPS cohort. Thresholds for genome-wide significance ($p=5x10^{-8}$, red line) and suggestive association ($p=1x10^{-5}$, blue line) are shown. The y-axis was truncated at $p=10^{-30}$.

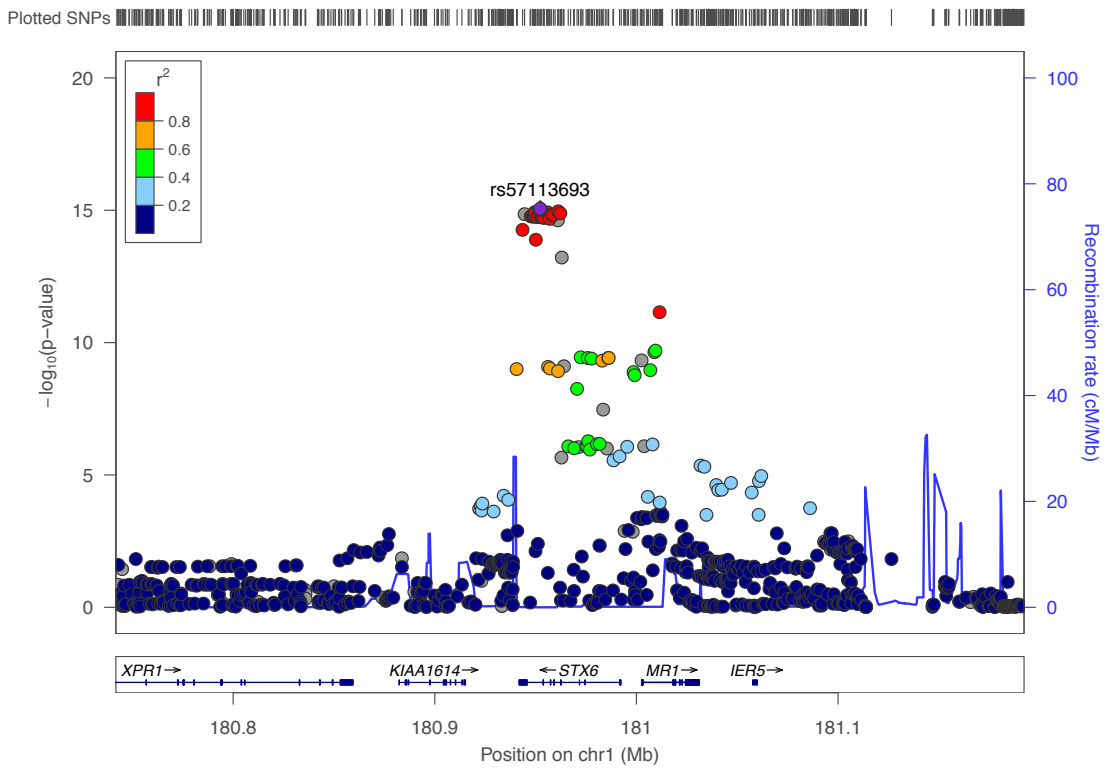
A


Plotted SNPs
rs1560310
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
EFTUD2 FMNL1 PLEKHM1 LOC644172 ARL17A WNT3
CCDC103 SPATA32 CRHR1 MAPT LRRC37A NSF GOSR2
GFAP MAP3K14−AS1 MGC57346 MAPT−IT1 KANSL1−AS1 NSFP1 WNT9B
KIF18B MAP3K14 CRHR1−IT1 STH ARL17B MIR5089
MIR6783 ARHGAP27 MAPT−AS1 KANSL1 LRRC37A2 RPRML
C1QL1 MIR4315−2 SPPL2C
DCAKD MIR4315−1
NMT1 LRRC37A4P
PLCD3
MIR6784
3 genes omitted
43 43.5 44 44.5 45
Position on chr17 (Mb)

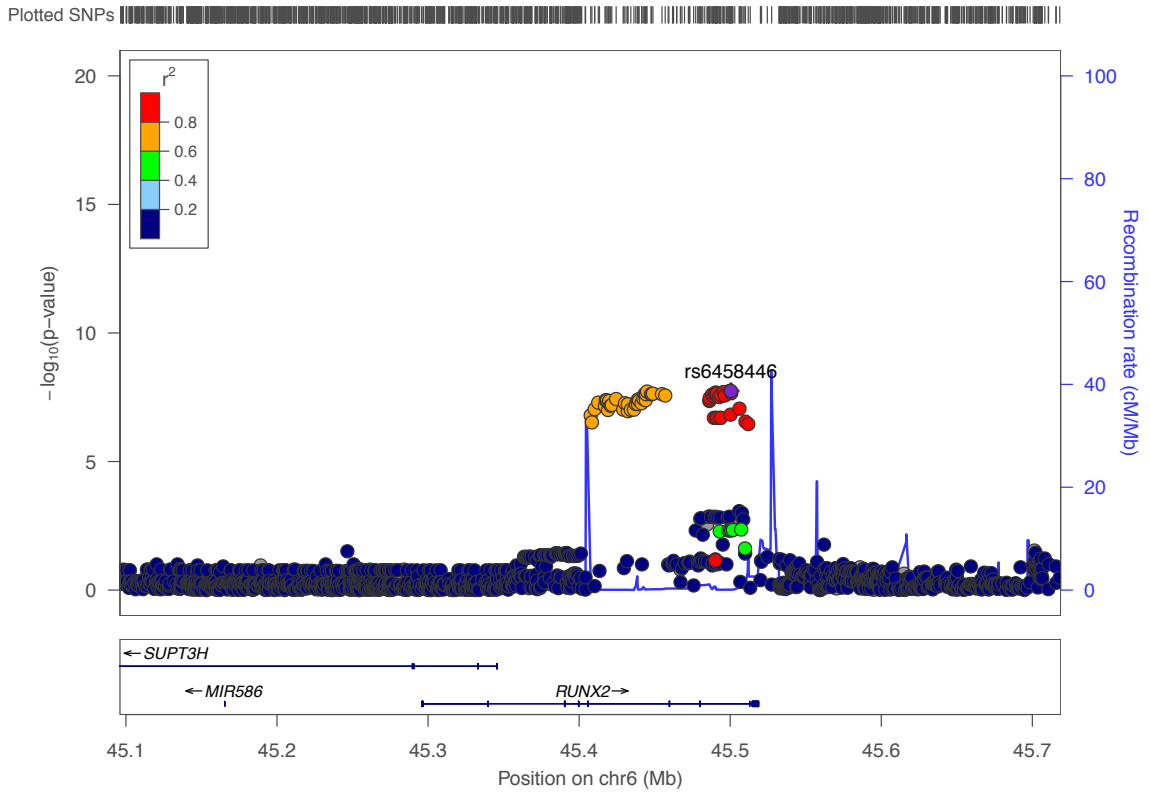
B


Plotted SNPs
rs10675541
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
CX3CR1 SLC25A38 MOBP
CCR8 RPSA
SNORA6
SNORA62
39.4 39.5 39.6 39.7
Position on chr3 (Mb)

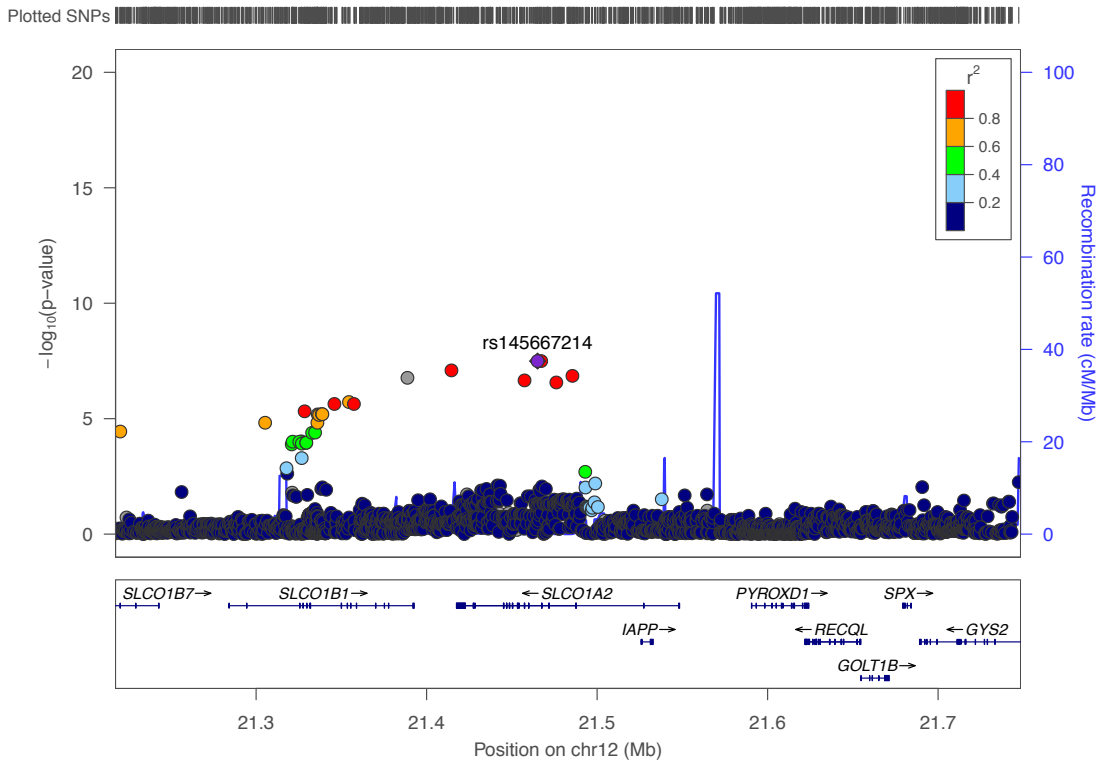
C


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs57113693
−log10(p−value)
Recombination rate (cM/Mb)
XPR1→
KIAA1614→
←STX6
MR1→
IER5→
180.8
180.9
181
181.1
Position on chr1 (Mb)

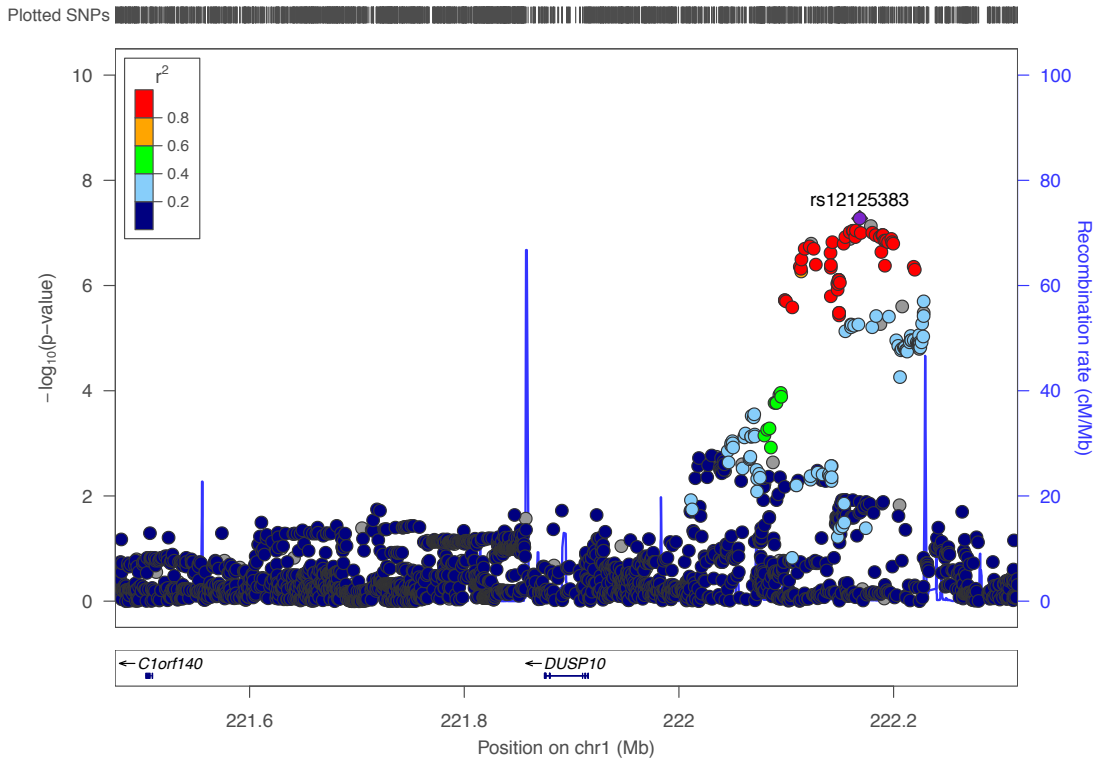
D


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs6458446
−log10(p−value)
Recombination rate (cM/Mb)
←SUPT3H
←MIR586
RUNX2→
45.1
45.2
45.3
45.4
45.5
45.6
45.7
Position on chr6 (Mb)

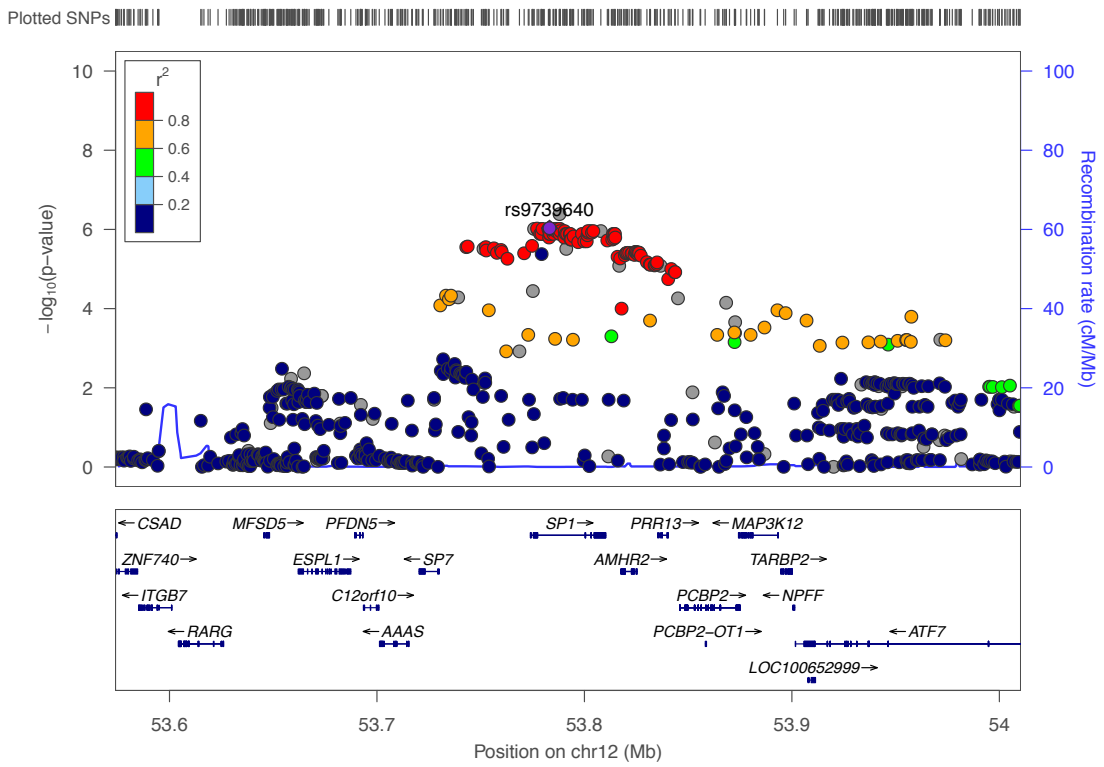
E


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs145667214
−log10(p−value)
Recombination rate (cM/Mb)
SLCO1B7→
SLCO1B1→
←SLCO1A2
PYROXD1→
SPX→
IAPP→
←RECQL
←GYS2
GOLT1B→
Position on chr12 (Mb)

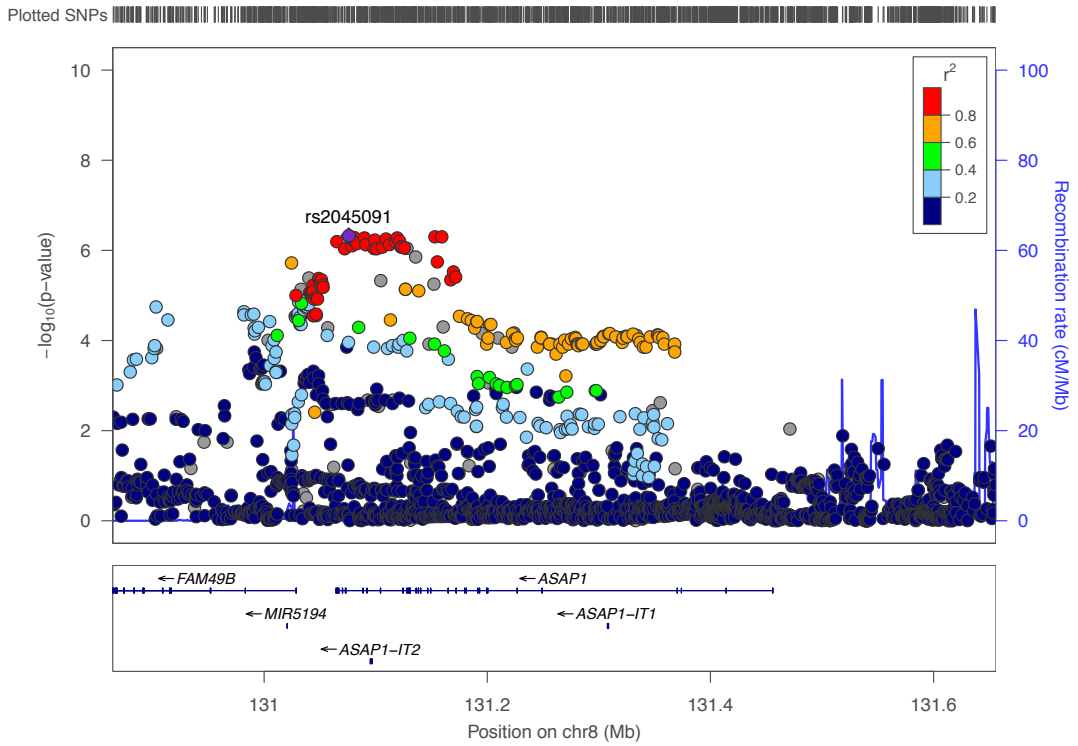
F


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs12125383
−log10(p−value)
Recombination rate (cM/Mb)
←C1orf140
←DUSP10
Position on chr1 (Mb)

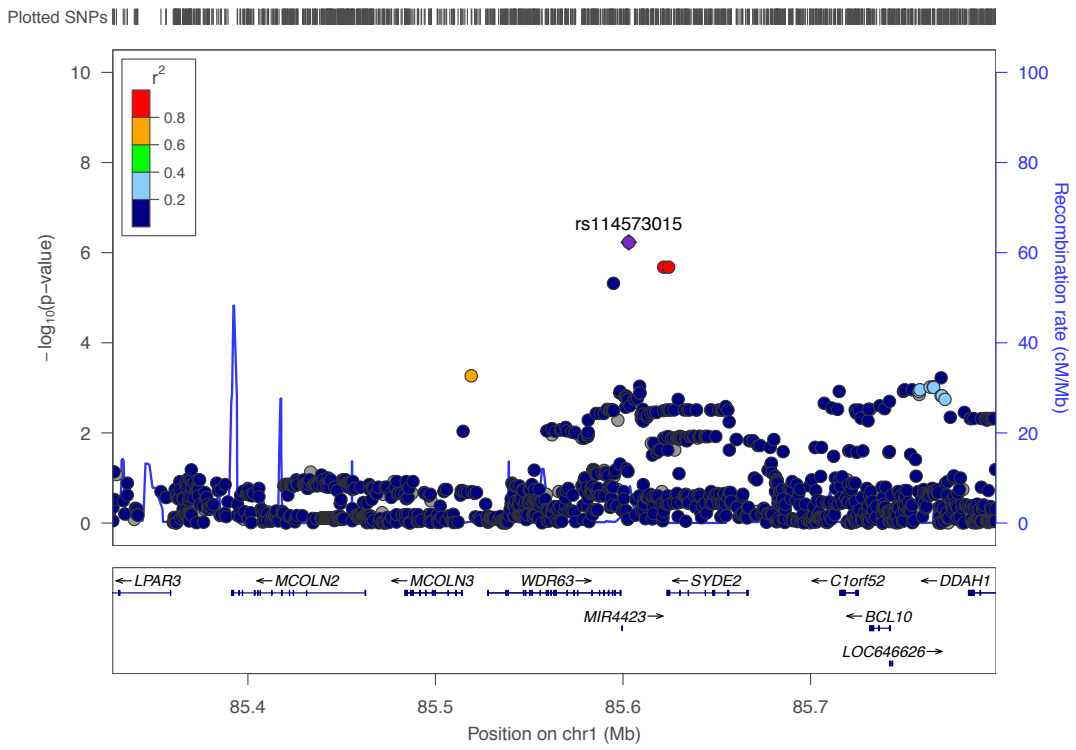
G


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs9739640
−log10(p−value)
Recombination rate (cM/Mb)
CSAD
MFSD5
PFDN5
SP1
PRR13
MAP3K12
ZNF740
ESPL1
SP7
AMHR2
TARBP2
ITGB7
C12orf10
PCBP2
NPFF
RARG
AAAS
PCBP2−OT1
ATF7
LOC100652999
Position on chr12 (Mb)

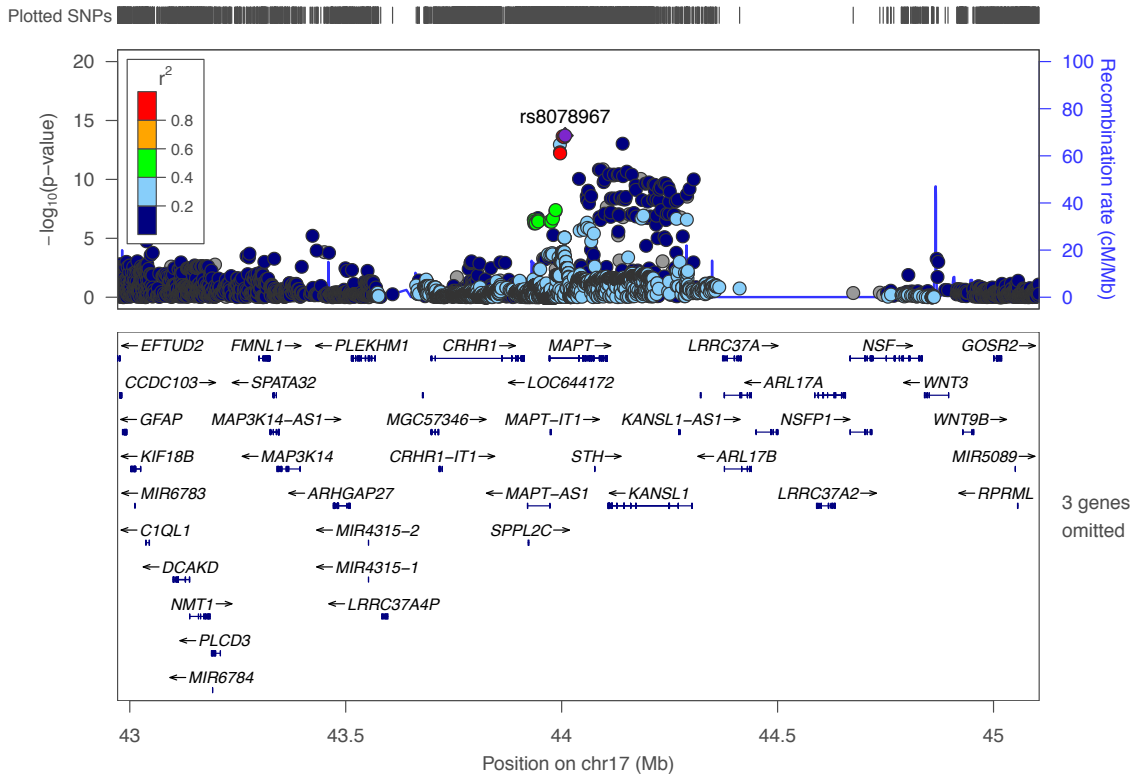
H


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs2045091
−log10(p−value)
Recombination rate (cM/Mb)
FAM49B
ASAP1
MIR5194
ASAP1−IT1
ASAP1−IT2
Position on chr8 (Mb)

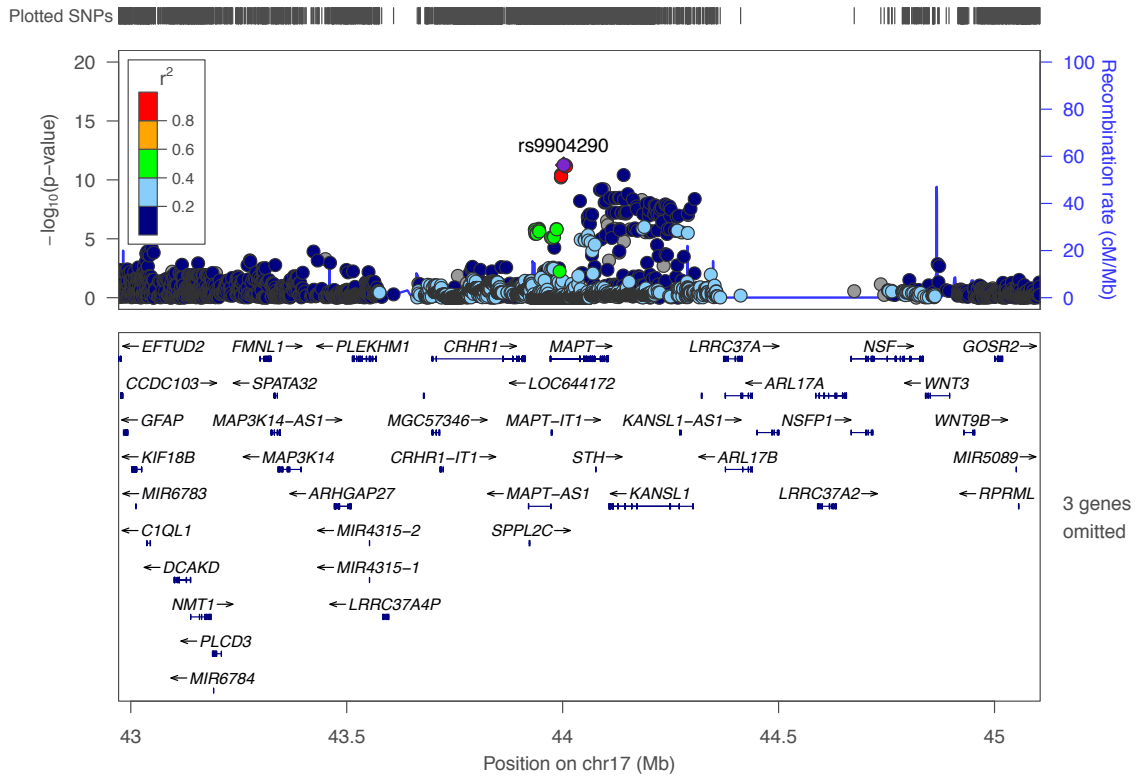
I


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs114573015
−log10(p−value)
Recombination rate (cM/Mb)
LPAR3
MCOLN2
MCOLN3
WDR63
SYDE2
C1orf52
DDAH1
MIR4423
BCL10
LOC646626
Position on chr1 (Mb)

J


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs8078967
−log10(p−value)
Recombination rate (cM/Mb)
EFTUD2
FMNL1
PLEKHM1
CRHR1
MAPT
LRRC37A
NSF
GOSR2
CCDC103
SPATA32
LOC644172
ARL17A
WNT3
GFAP
MAP3K14−AS1
MGC57346
MAPT−IT1
KANSL1−AS1
NSFP1
WNT9B
KIF18B
MAP3K14
CRHR1−IT1
STH
ARL17B
MIR5089
MIR6783
ARHGAP27
MAPT−AS1
KANSL1
LRRC37A2
RPRML
C1QL1
MIR4315−2
SPPL2C
DCAKD
MIR4315−1
NMT1
LRRC37A4P
PLCD3
MIR6784
3 genes omitted
Position on chr17 (Mb)

K



Supplementary Figure A-6: Local association results at genome-wide significant and suggestive loci: A) 17q21.31, near MAPT; B) 3p22.1, near *MOBP*; C) 1q25.3, near STX6; D) 6p21.1, near *RUNX2*; E) 12p12.1, near *SLCO1A2*; F) 1q41, near *DUSP10* in an intergenic region; G) 12q13.13, near *SP1*; H) 8q24.21, near *ASAP1*; and I) 1p22.3, near *WDR63*. J) The association at 17q21.31 corresponded with an extended haplotype with two non-recombining alleles, H1 and H2. Following regression of the haplotype status, haplotype-independent regional association was found in 17q21.31. K) Similar results are found considering only H1/H1 individuals. Representative SNPs are highlighted. The plots were made with LocusZoom (http://locuszoom.sph.umich.edu/locuszoom/).

Supplementary Table A-1: Characteristics of the individual study cohorts.

Cohort	Study	Array	Genotyped SNPs	Filtered Patients	Filtered Controls	Unfiltered Patients
Joint Analysis	TOTAL	Combined (Imputed)	5,523,934	1,646	11,619	13,334
	UCLA-Omni2.5 (Imputed)	Illumina HumanOmni 2.5	8,494,997	249	1050	
	UCLA-HumanCore (Imputed)	Illumina HumanCore/Omni Express	6,754,310	66	888	
	NNIPPS (Imputed)	Illumina HumanOmni 2.5	8,530,877	308	1,330	
	Hoglinger (Imputed)	Illumina HumanQuad 660W	7,634,768	1,069	8,374	
UCLA-Omni 2.5	UCLA PSP and Controls - Illumina	Illumina HumanOmni 2.5	2,197,833	271	89	399
	*phs000371.v1.p1 "Genetic Modifiers of Huntington's Disease"	Illumina HumanOmni 2.5	2,379,855	0	120	1,233
	*phs000429.v1.p1 "NEI Age-Related Eye Disease Study (AREDS) - Genetic Variation in Refractive Error Substudy"	Illumina HumanOmni 2.5	2,182,680	0	678	1,667
	*phs000421.v1.p1 "A Genome-Wide Association Study of Fuchs' Endothelial Corneal Dystrophy (FECD)"	Illumina HumanOmni 2.5	2,443,177	0	199	3,537
	* the NNIPPS and UCLAOmni2.5 groupings include non-overlapping controls from a shared cohort of combined Omni2.5 studies					
UCLA-HumanCore	UCLA PSP and Controls - NYGC	Illumina HumanCore	547,644	78	41	221
	WTCCC2 Controls	Illumina OmniExpress	730,525	0	870	870
NNIPPS	NNIPPS study PSP - Ammar Al Chalabi	Illumina HumanOmni 2.5	2,379,855	341	0	764
	*phs000371.v1.p1 "Genetic Modifiers of Huntington's Disease"	Illumina HumanOmni 2.5	2,379,855	0	155	1,233
	*phs000429.v1.p1 "NEI Age-Related Eye Disease Study (AREDS) - Genetic Variation in Refractive Error Substudy"	Illumina HumanOmni 2.5	2,182,680	0	989	1,667
	*phs000421.v1.p1 "A Genome-Wide Association Study of Fuchs' Endothelial Corneal Dystrophy (FECD)"	Illumina HumanOmni 2.5	2,443,177	0	223	3,537
	* the NNIPPS and UCLAOmni2.5 groupings include non-overlapping controls from a shared cohort of combined Omni2.5 studies					
Hoglinger	Hoglinger et al. Cases Only from NIAGADS	Illumina HumanQuad 660W	561,882	1,112	0	1,112
	phs000103.v1.p1 "Genome-Wide Association Studies of Prematurity and Its Complications"	Illumina HumanQuad 660W	592,839	0	974	3,945
	phs000289.v1.p1 "National Human Genome Research	Illumina HumanQuad 660W	657,366	0	1,326	2,760

Institute (NHGRI) GENEVA Genome-Wide Association Study of Venous Thrombosis"					
phs000188.v1.p1 "Vanderbilt Genome-Electronic Records (VGER) Project: QRS Duration"	Illumina HumanQuad 660W	543,515	0	2,403	2,403
phs000203.v1.p1 "A Genome-Wide Association Study of Peripheral Arterial Disease"	Illumina HumanQuad 660W	657,366	0	1,666	3,432
phs000237.v1.p1 "Northwestern NUgene Project: Type 2 Diabetes"	Illumina HumanQuad 660W	657,366	0	239	607
phs000243.v1.p1 "Group Health/UW Aging and Dementia eMERGE study"	Illumina HumanQuad 660W	657,366	0	884	1,618
phs000170.v1.p1 "A Genome-Wide Association Study on Cataract and HDL in the Personalized Medicine Research Project Cohort"	Illumina HumanQuad 660W	560,635	0	1,264	3,947

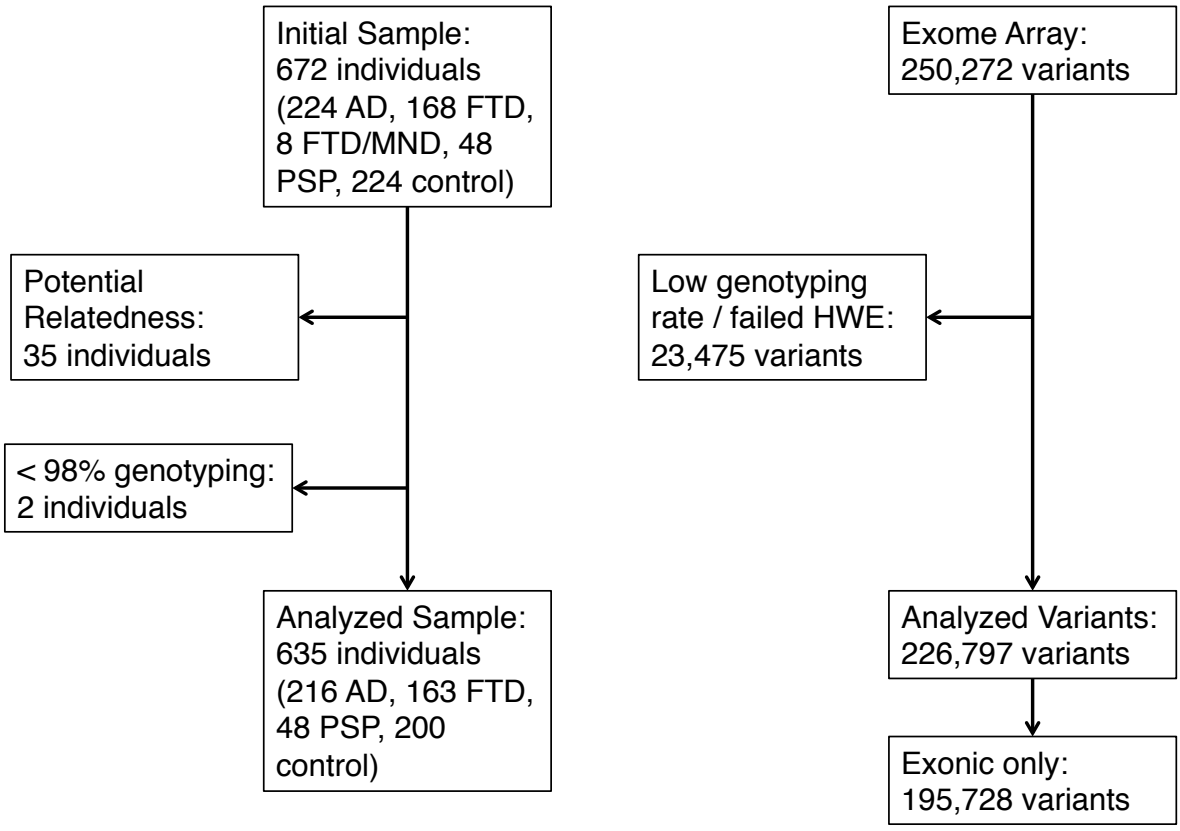
Supplementary Table A-2: CNV burden in PSP cases vs. controls, stratified by CNV type (deletion vs. duplications) and CNV size.

		Average CNV Count per Subject			Average Number of Bases Spanned per CNV (KB)			Average Number of Genes Spanned per Subject			Average Genes per Kilobase CNV		
Copy Number	**CNV Size**	**Cases**	**Controls**	**P-value**	**Cases**	**Controls**	**P-value**	**Cases**	**Controls**	**P-value**	**Cases**	**Controls**	**P-value**
Deletion	ALL	0.986	0.856	0.125	170.700	163.200	0.313	0.292	0.292	0.511	0.003	0.003	0.904
Deletion	50k - 200k	0.804	0.693	0.136	98.300	101.300	0.800	0.260	0.257	0.479	0.004	0.004	0.799
Deletion	200k - 500k	0.160	0.150	0.387	293.600	291.700	0.439	0.028	0.032	0.683	0.001	0.001	0.743
Deletion	500k - 1M	0.036	0.021	0.122	633.600	623.600	0.413	0.004	0.006	0.842	0.000	0.001	0.831
Deletion	1M+	0.007	0.010	0.779	2040.000	1363.000	0.040	0.000	0.001	1.000	0.000	0.000	0.634
Duplication	ALL	1.292	1.108	0.106	226.600	248.100	0.842	0.409	0.344	0.176	0.003	0.002	0.294
Duplication	50k - 200k	0.932	0.769	0.106	107.100	100.200	0.033	0.367	0.289	0.120	0.004	0.004	0.407
Duplication	200k - 500k	0.342	0.257	0.019	336.700	318.900	0.064	0.039	0.043	0.663	0.000	0.001	0.922
Duplication	500k - 1M	0.071	0.088	0.819	662.500	660.300	0.463	0.004	0.010	0.939	0.000	0.000	0.840
Duplication	1M+	0.014	0.027	0.927	1407.000	1575.000	0.704	0.000	0.002	1.000	0.000	0.000	0.511

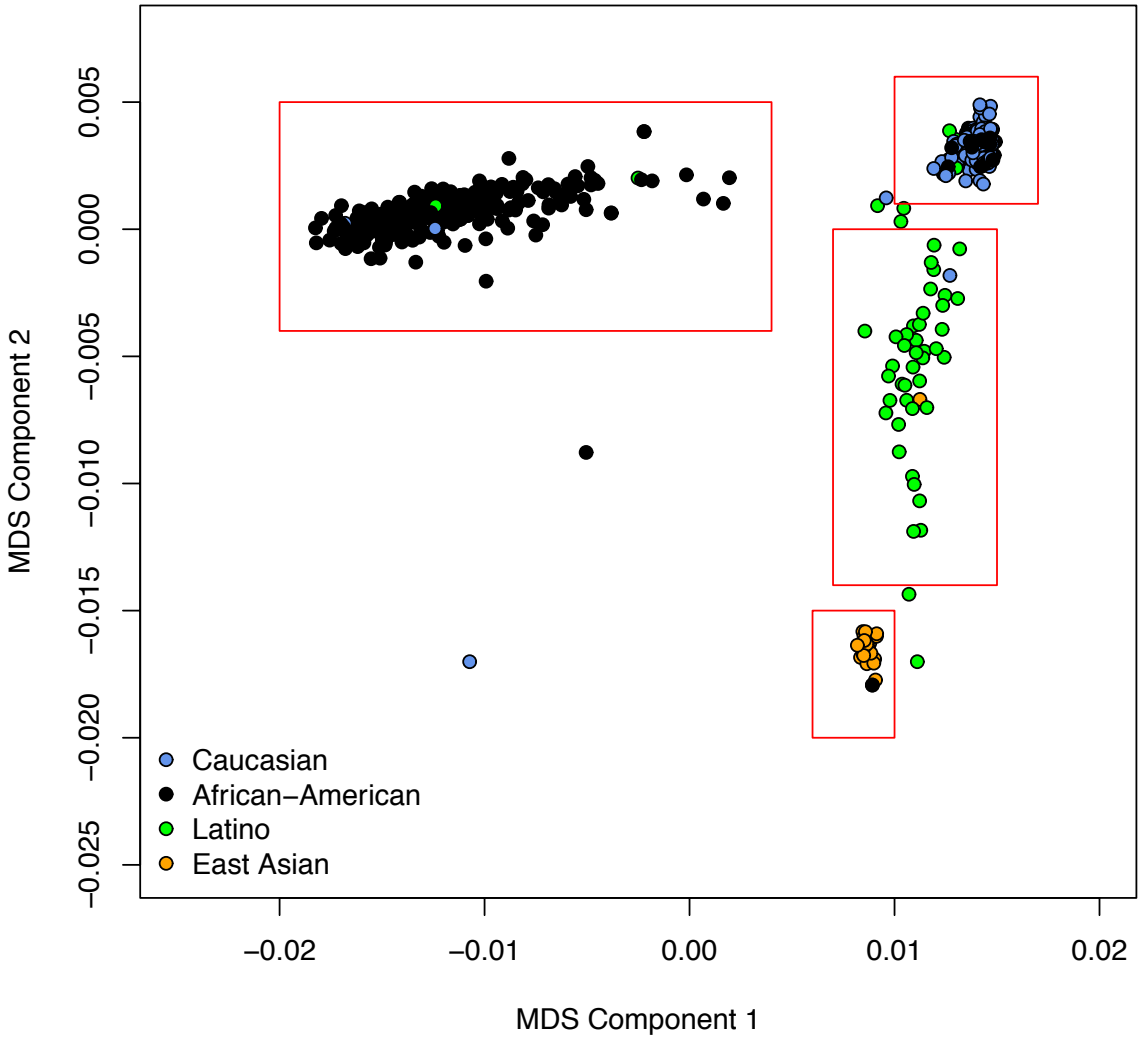
Supplementary Table A-3: Variance explained by significant and suggestive SNPs.

SNP	Gene	Previously Reported	Minor Allele	Odds Ratio (minor allele)	Minor Allele Frequency	Variance in Liability Explained
rs71920662	MAPT	Reported	T	0.19	0.22	0.05
rs10675541	MOBP	Reported	TAC	0.71	0.49	0.0035
rs57113693	STX6	Reported	C	1.35	0.4	0.0026
rs35740963	RUNX2	Novel	insT	0.77	0.35	0.0019
rs7966334	SLCO1A2	Novel	C	1.5	0.058	0.001
Suggestive						
rs12125383	DUSP10	Novel	A	1.28	0.19	0.0011
rs147124286	SP1	Novel	GA	0.74	0.18	0.0016
rs2045091	ASAP1	Novel	T	1.25	0.17	0.00086
rs114573015	WDR63	Novel	G	2.08	0.01	0.00067

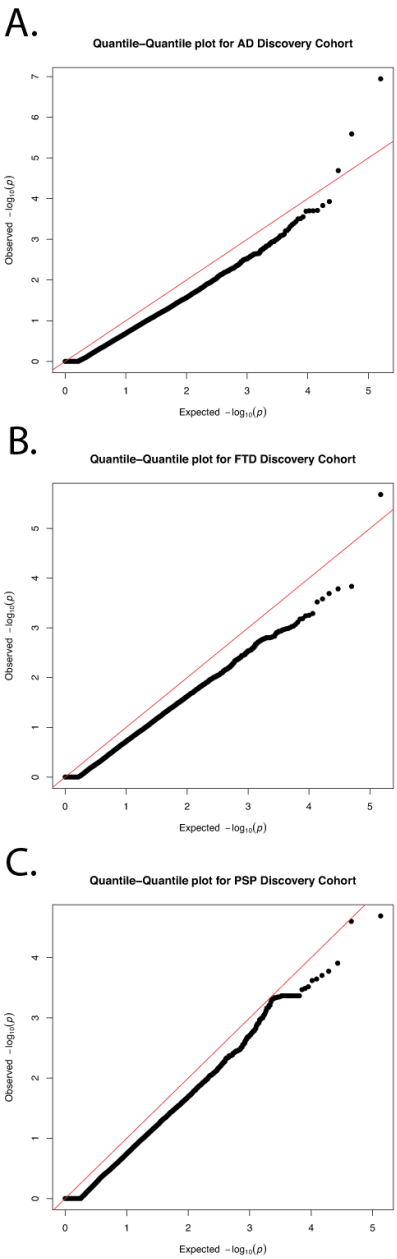
Appendix B: Supplementary Material for Chapter 3



Supplementary Figure B-1: Flowchart for quality control procedures for the GIFT cohort and exome array variants.



Supplementary Figure B-2: Scatterplot demonstrating the population structure within the replication cohort evidenced from multidimensional scaling (MDS). The first two MDS components are shown. Clusters were defined manually using the displayed boxes; subjects with reported ethnicity that differed from the majority of the cluster were removed from further analysis.



Supplementary Figure B-3: Quantile-quantile plots for the variant-level association statistics calculated from the discovery cohort for a) Alzheimer's disease, b) frontotemporal dementia, and c) progressive supranuclear palsy.

Supplementary Table B-1: Replication of associations for progressive supranuclear palsy genome-wide association studies.

SNP	Chr.	hg19 Coord.	Mapped Gene	Risk Allele	Lit. Risk Allele	Reference	Minor Allele Odds Ratio	p-value
rs1411478	1	180962282	*STX6*	A	A	Hoglinger et al., 2011	1.416	0.1473
rs6687758	1	222164948	Intergenic	G	G	Hoglinger et al., 2011	1.26	0.407
rs6547705	2	87044316	*CD8B*	A	A	Hoglinger et al., 2011	0.8465	0.556
rs7571971	2	88895351	*EIF2AK3*	T	T	Hoglinger et al., 2011	1.407	0.1633
rs1768208	3	39523003	*MOBP*	T	T	Hoglinger et al., 2011	1.616	0.04625
rs6852535	4	123478716	*IL2/IL21*	A*	G	Hoglinger et al., 2011	1.012	0.9604
rs12203592	6	396321	*IRF4*	C	C	Hoglinger et al., 2011	0.849	0.5826
rs242557	17	44019712	*MAPT*	A	A	Hoglinger et al., 2011	1.38	0.1585
rs8070723	17	44081064	*MAPT*	A	A	Hoglinger et al., 2011	0.1796	0.000431

* does not match literature risk allele

Supplementary Table B-2: Gene-level association statistics for Alzheimer's disease using the Sequence Kernel Association Test (SKAT) for genes with FDR < 50% in the discovery cohort.

Gene	No. Typed Variants	Disc. AD p-value	Disc. FTD p-value	Disc. PSP p-value	Rep. Caucasian p-value	Rep. African-American p-value	Rep. Latino p-value	Rep. Asian p-value	HGMD Alzheimer's Disease Gene
DYSF	84	$5.5x10^{-5}$	0.15	0.22	0.076	0.45	0.10	1	FALSE
PAXIP1	7	0.00023	0.21	0.43	0.30	0.81	0.016	0.037	FALSE
TOP1MT	33	0.00029	0.0062	1	0.31	0.0060	0.40	0.40	FALSE
C3orf1	8	0.00039	0.36	0.82	0.70	0.10	0.56	0.81	FALSE
SETDB1	14	0.00041	0.78	0.18	0.48	0.30	1	0.23	FALSE
CRISPLD1	12	0.00045	0.23	0.081	0.71	1	1	0.66	FALSE

Supplementary Methods

Subject recruitment and diagnosis. Patients are referred to these research studies primary through referrals from clinicians in the community. All participants in these project underwent a standard multidisciplinary diagnostic assessment including neurological history and examination, nursing assessment, laboratory evaluation, and a previously described neuropsychological assessment of memory, executive function, visuospatial ability, language, and mood (Kramer et al., 2003). AD patients were diagnosed by applying NINCDS-ADRDA criteria for probable Alzheimer's disease based on neurological and neuropsychological examination, brain imaging and laboratory assessments to rule out other causes of dementia. All subjects and/or their proxies signed informed consents for genetic studies.

Genotyping. Polymorphisms at APOE (rs429358 and rs7412) were genotyped using the pre-designed TaqMan genotyping assays; a polymorphism tagging MAPT H1/H2 (rs1560310) was genotyped using a custom TaqMan genotyping assay. Exome array genotyping using the Illumina HumanExome arrays was performed by the UCLA Neuroscience Genomics Core using an Illumina iScan confocal laser scanner. For the discovery cohort, samples were randomly assigned to arrays (12 samples per array) and genotyped at the same time. For the replication cohort, samples were randomly assigned to arrays (12 samples per array) and genotyped at the same time, separate from the discovery cohort.

Data pre-processing. First, known and cryptically related individuals were removed. Cryptic relatedness was determined by IBD estimation in PLINK version 1.07 (Purcell et al., 2007). The set of exome array variants was pruned to 18,250 SNPs in approximate linkage equilibrium based on pairwise genotypic correlation, using a window of 50 SNPs, a step size of 5 SNPs, and an r^2 threshold of 0.5. A total of 32 samples with either self-reported familial relationships or with proportion of alleles IBD greater than 0.2 were removed from further analysis. Second,

23,475 variants with lower than a 98% genotyping rate (21,729 in total), or not in Hardy-Weinberg equilibrium (a *p*-value threshold of 10^{-4}; 2,308 in total) were excluded. Following quality control procedures, the initial discovery cohort was comprised of 226,797 variants genotyped in 216 patients with AD, 163 patients with FTD, 48 patients with PSP, and 200 non-demented controls. Of these 627 patients, analysis of X-chromosome homozygosity using PLINK (--check-sex) identified that gender was consistent in 621 (99.0%) of them (3 cases in which gender could not be called, and 3 cases with discordant gender call).

Heritability estimation. The variance explained by the subset of variants was determined using GCTA version 1.13 (Yang et al., 2011). A genetic relationship matrix (GRM) was computed for all of the subjects, using 1) all of the typed variants, 2) the exonic content of the chip, or 3) the exonic content of the chip with lower than 5% minor allele frequency within our total cohort. A principal components analysis was performed on the GRM using GCTA. Then, GCTA's REML analysis was used to determine the variance explained by each set of variants, using the first four principal components as a covariate to correct for genetic ancestry.

Variant-level association testing. Variant-level association testing was performed using logistic regression implemented in PLINK (Purcell et al., 2007). In order to correct for population stratification, multidimensional scaling was performed within PLINK to extract the first four multidimensional scaling axes. These four dimensions were then taken as covariates in the logistic model. In order to confirm the validity of this approach, association was also performed using FaST-LMM, a linear mixed model method (Lippert et al., 2011) that has been shown to correct for test statistic inflation in the presence of population structure confounders and cryptic relatedness. To avoid problems of "proximal contamination", a subset of 18,875 polymorphisms with minor allele frequency (MAF) greater than 5%, genotyping rate greater than 99%, and in approximate linkage equilibrium based on pairwise genotypic correlation, using a window of 50

SNPs, a step size of 5 SNPs, and an r^2 threshold of 0.5 was used to estimate the genetic similarity matrix. A suggestive p-value threshold of $1x10^{-5}$ was used for the initial screening stage, as described for previous association studies (Duggal et al., 2008). Power calculations mentioned in the text were performed using the QUANTO version 1.2.4 software (University of Southern California, Los Angeles, CA), assuming a log-additive genetic model. For power calculations and GCTA modeling, population prevalence of neurodegenerative disease in elderly individuals was estimated at 11% in AD (Hebert et al., 2013), 0.022% in FTD (Borroni et al., 2010), and 0.008% in PSP (Bower et al., 1997).

Gene-level association testing. Gene-level association testing was performed using the Sequence Kernel Association Test (SKAT) (Wu et al., 2011) and implemented in the R package 'SKAT', R version 2.15.1 (R Foundation for Statistical Computing, Vienna, Austria). To control for population structure, the first four principal components of the genotyping data were input as covariates into the SKAT program. The designed exonic content of the exome array, comprising a subset of 195,728 non-synonymous variants (missense and nonsense) and splice variants, were selected from the quality-controlled exome array content for gene-level testing. A false discovery rate (FDR) was estimated by generating *B*=100 permutations of the genotyping data by randomly shuffling the case and control status of each subject and performing the SKAT analysis, as previously described (Xie et al., 2005). Briefly, each SKAT p-value was tested as a potential threshold *d* for the test statistic. An estimate of the FDR controlled at each p-value threshold *d* was calculated as follows:

$$\mathrm{FDR}(d) = \frac{FP(d)}{TP(d)}(1 - S)$$

Here, *FP*(*d*) is an estimate of the number of false positive genes at the threshold *d*, given by

$$FP(d) = \sum_{b=1}^{B} \sum_{i} I(p_{i,b} \leq d)/B$$

where *B* is the total number of permutations (100), *I* is the indicator function, and $p_{i,b}$ is the p-value of the *i*th gene in the *b*th permutation. Similarly, *TP*(*d*) is an estimate of the total number of positive genes (sum of true and false positives), calculated by

$$TP(d) = \sum_{i} I(p_i \leq d)$$

where p_i is the p-value of the *i*th gene in the experimental dataset. The number of true positive genes (*S*) was conservatively assumed to be vanishingly small compared to the total number of genes. We selected an FDR threshold of 50% to prioritize genes for analysis in a follow-up replication cohort for Alzheimer's disease, and an FDR threshold of 15% for suggestive results for FTD and PSP (for which additional patients were not available). For the second stage of testing, the p-value threshold for a single gene was determined using Bonferroni correction, where the number of hypothesis tests was calculated as the product of the total number of genes that passed the first stage of testing and the number of ethnicities tested in the sample (4), at a family-wise error rate of 0.05.

Expression analysis of publically available data. The expression of mRNA in the brains of patients with Alzheimer's disease and non-demented controls was described by Zhang et al. (2013), who measured gene expression (in each of prefrontal cortex, visual cortex, and cerebellum) in 415 cases and 171 controls using microarray. Data presented in the Zhang et al. manuscript was accessed from the Sage Bionetworks Synapse service (ID syn4505).

Expression data corresponding to *DYSF* (NM_003494) or *PAXIP1* (NM_007349) was extracted. The expression data had been corrected for technical covariates such as age, gender, post-mortem interval, RNA integrity, and others. Differential expression between patients with Alzheimer's disease and non-demented controls was assessed using the Welch two-sample t-test implemented in R. An additional dataset of microarray expression data described by Webster et al. (2009) from various brain regions (frontal, temporal, parietal, and cerebellar cortices) of 176 Alzheimer's disease patients and 188 controls was similarly analyzed. Data presented in the Webster et al. manuscript was accessed from the NCBI Gene Expression Omnibus (GEO Accession GSE15222). This dataset contained a probe for *DYSF* expression, but did not measure the expression of *PAXIP1*. Expression data had been adjusted for technical covariates such as gender, age, APOE status, post-mortem interval, brain region, and others, and was rank invariant normalized.

References

Borroni, B., Alberici, A., Grassi, M., Turla, M., Zanetti, O., Bianchetti, A., Volta, G.D., Rozzini, R., Gilberti, N., Bellelli, G.*, et al.* (2010). Is Frontotemporal Lobar Degeneration a Rare Disorder? Evidence from a Preliminary Study in Brescia County, Italy. Journal of Alzheimer's Disease *19*, 111-116.

Bower, J.H., Maraganore, D.M., McDonnell, S.K., and Rocca, W.A. (1997). Incidence of progressive supranuclear palsy and multiple system atrophy in Olmsted County, Minnesota, 1976 to 1990. Neurology *49*, 1284-1288.

Duggal, P., Gillanders, E., Holmes, T., and Bailey-Wilson, J. (2008). Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics *9*, 516.

Hebert, L.E., Weuve, J., Scherr, P.A., and Evans, D.A. (2013). Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology *80*, 1778-1783.

Kramer, J.H., Jurik, J., Sha, S.J., Rankin, K.P., Rosen, H.J., Johnson, J.K., and Miller, B.L. (2003). Distinctive Neuropsychological Patterns in Frontotemporal Dementia, Semantic Dementia, And Alzheimer Disease. Cognitive and Behavioral Neurology *16*, 211-218.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. Nature Methods *8*, 833-835.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J.*, et al.* (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. American Journal of Human Genetics *81*, 559-575.

Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M.*, et al.* (2009). Genetic Control of Human Brain Transcript Expression in Alzheimer Disease. American Journal of Human Genetics *84*, 445-458.

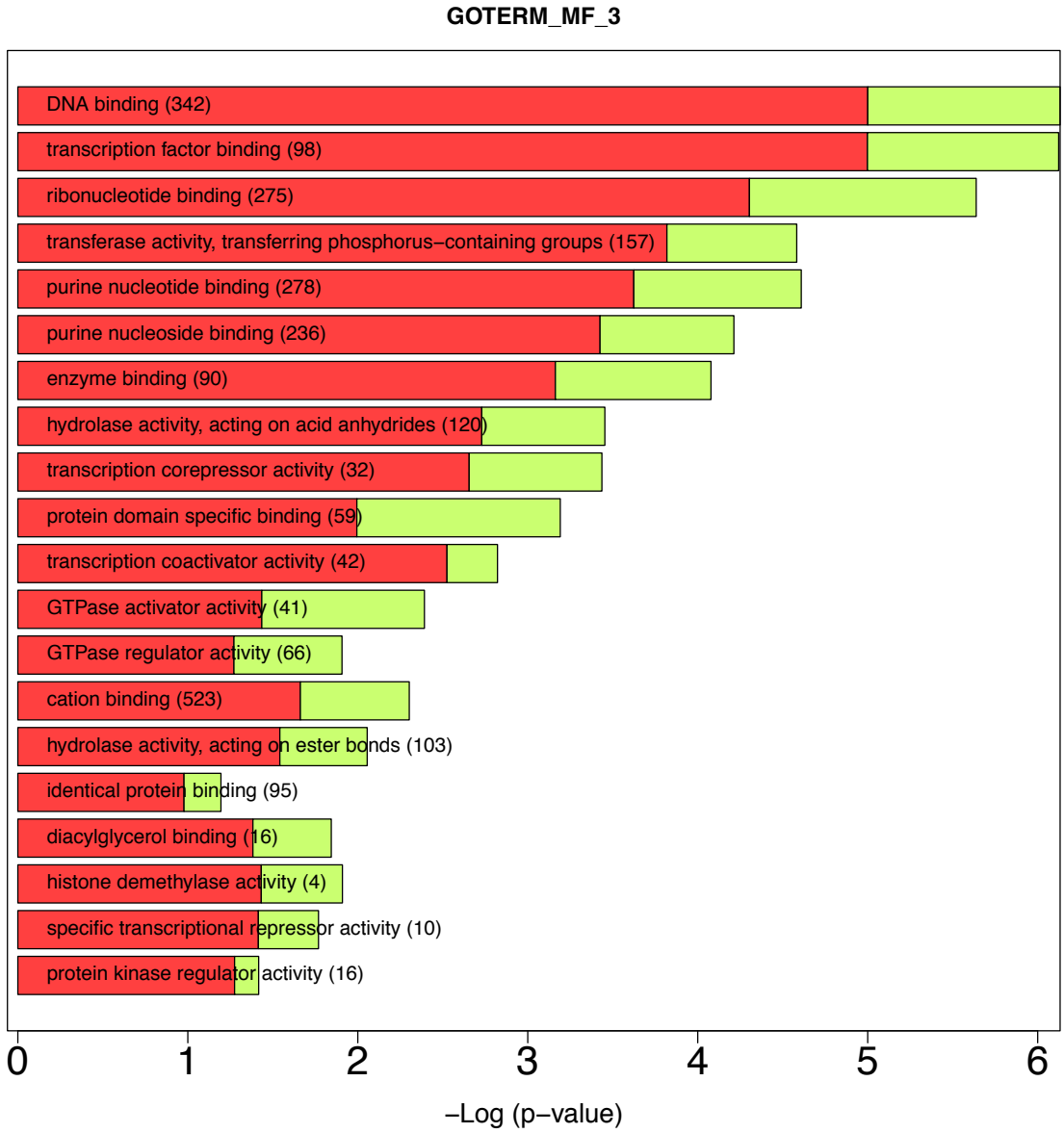
Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. American Journal of Human Genetics *89*, 82-93.

Xie, Y., Pan, W., and Khodursky, A.B. (2005). A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. Bioinformatics *21*, 4280-4288.

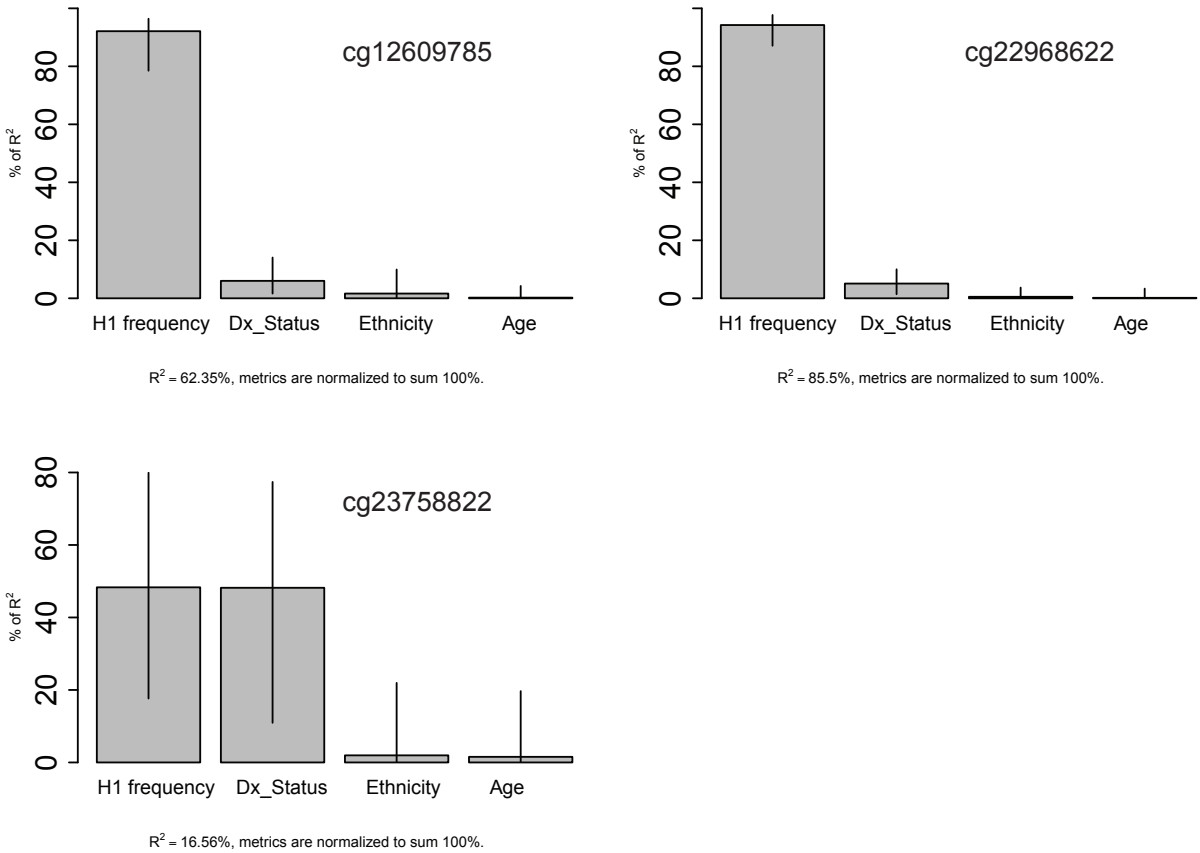
Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. American Journal of Human Genetics *88*, 76-82.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, Alexei A., Zhang, C., Xie, T., Tran, L., Dobrin, R.*, et al.* (2013). Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. Cell *153*, 707-720.

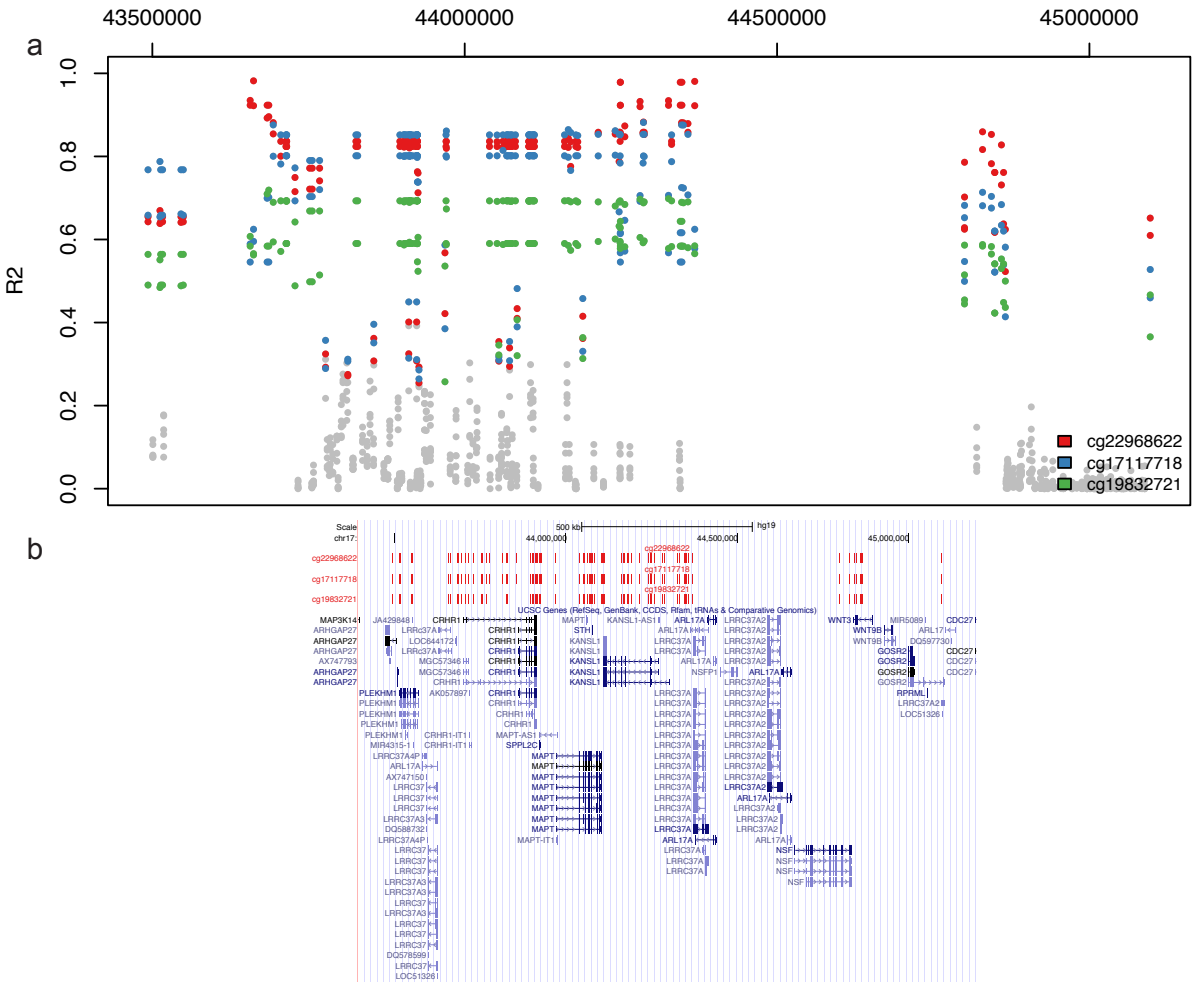
Appendix C: Supplementary Material for Chapter 5



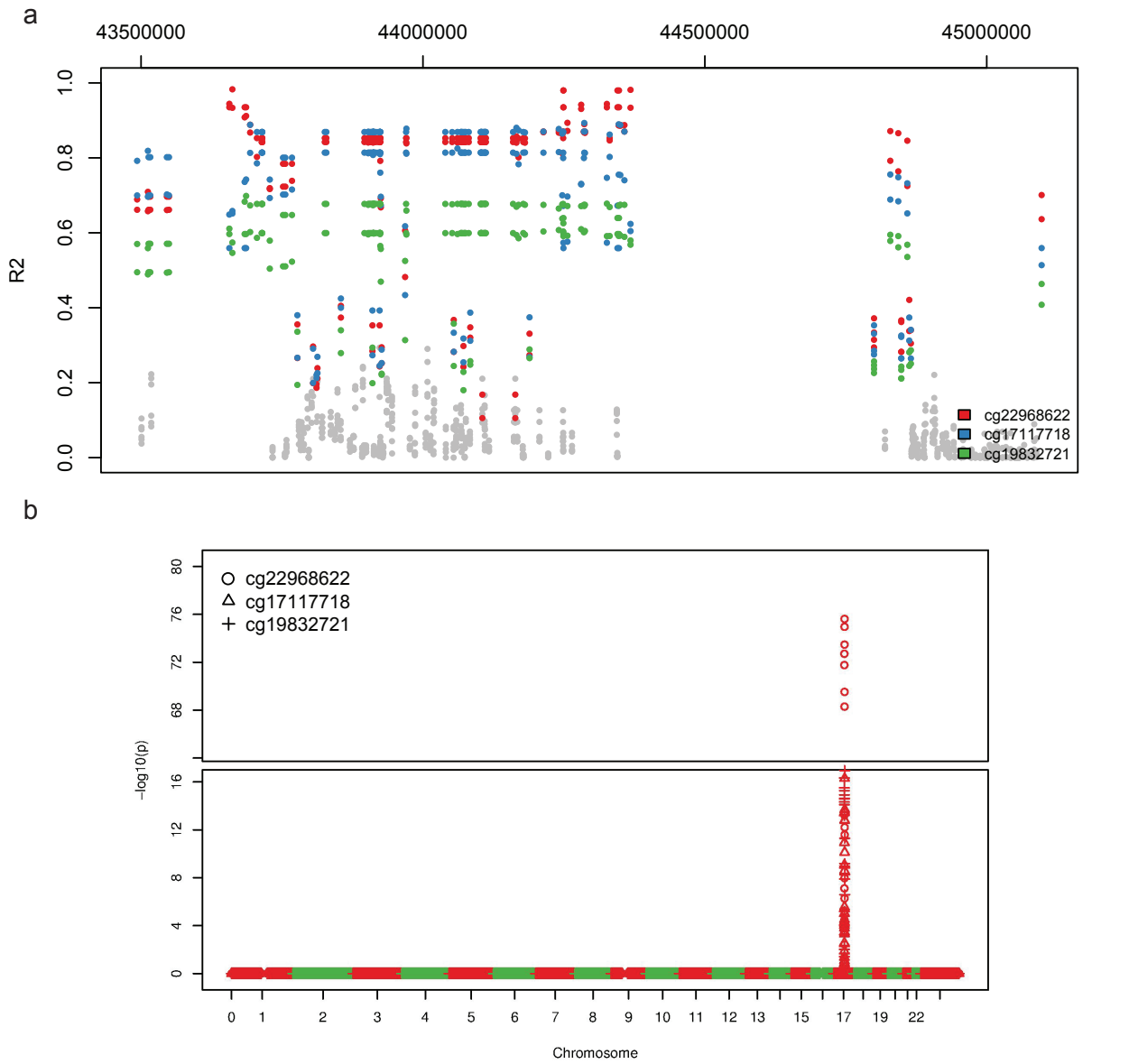
Supplementary Figure C-1: Over-represented gene ontology (GO), molecular function, level 3 (MF_3) categories among DMPs in PSP versus controls (in green the proportion of hypomethylated DMPs; in red the proportion of hypermethylated DMPs) sorted by −log10 (p-value). A −log (p-value) of 1.3 corresponds to an over-representation *p*-value of 0.05.



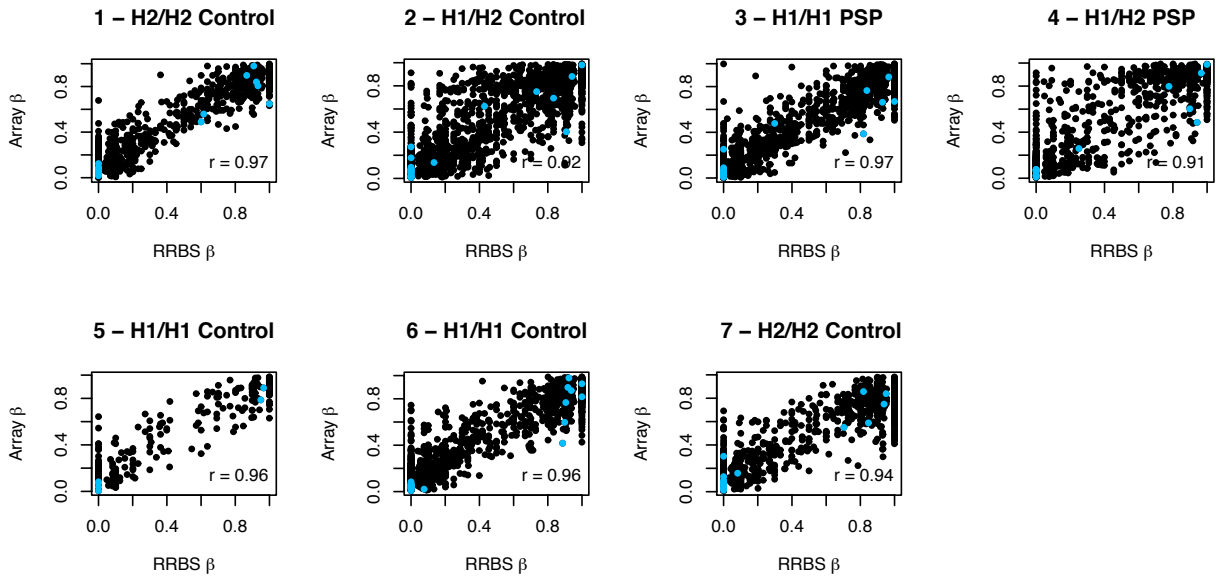
Supplementary Figure C-2: Relative effect of covariates on methylation beta value variance. The H1 haplotype accounts for most of the explained variance. For the top 3 PSP-related DMPs, the relative importance of predictors in the multivariate linear regression model (including H1 frequency, diagnosis status, age, and ethnicity) was calculated using R package relaimpo. Error bars: 95% bootstrap confidence intervals.



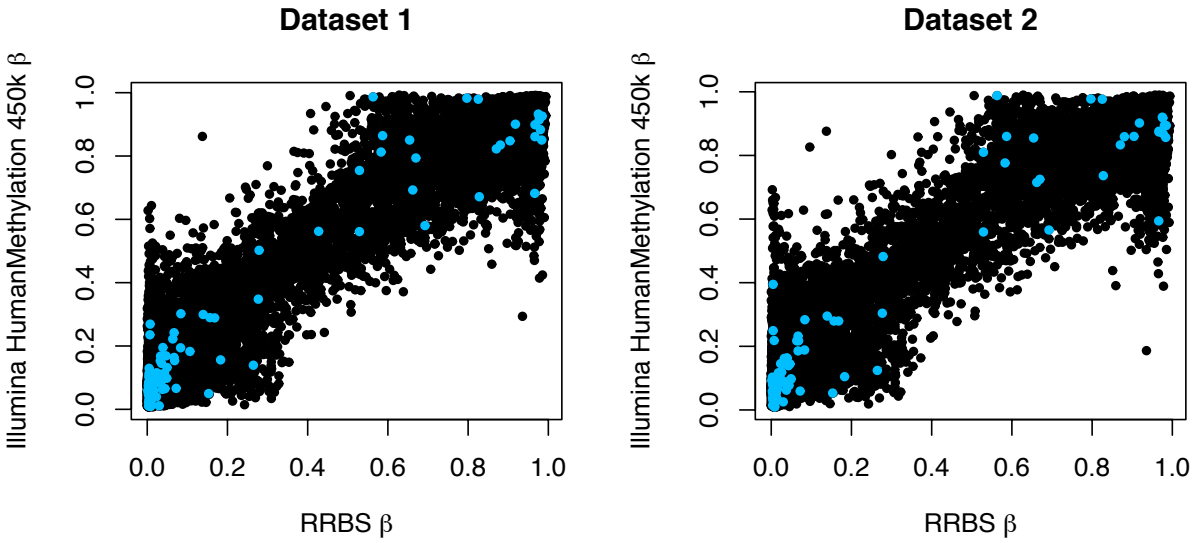
Supplementary Figure C-3: Methylation QTL analysis in 226 individuals of European descent. (a) scatterplot representing the R-squared for each of the SNPs in the 17q21.31 region associated with at least one of the three DMPs. Gray: not significant SNPs. (b) corresponding genomic region at 17q21.31 (UCSC Genome Browser, hg19). Top significant SNPs controlling the three DMPs are highlighted in red. Age and the first two principal components generated from MDS analysis were added as covariates.



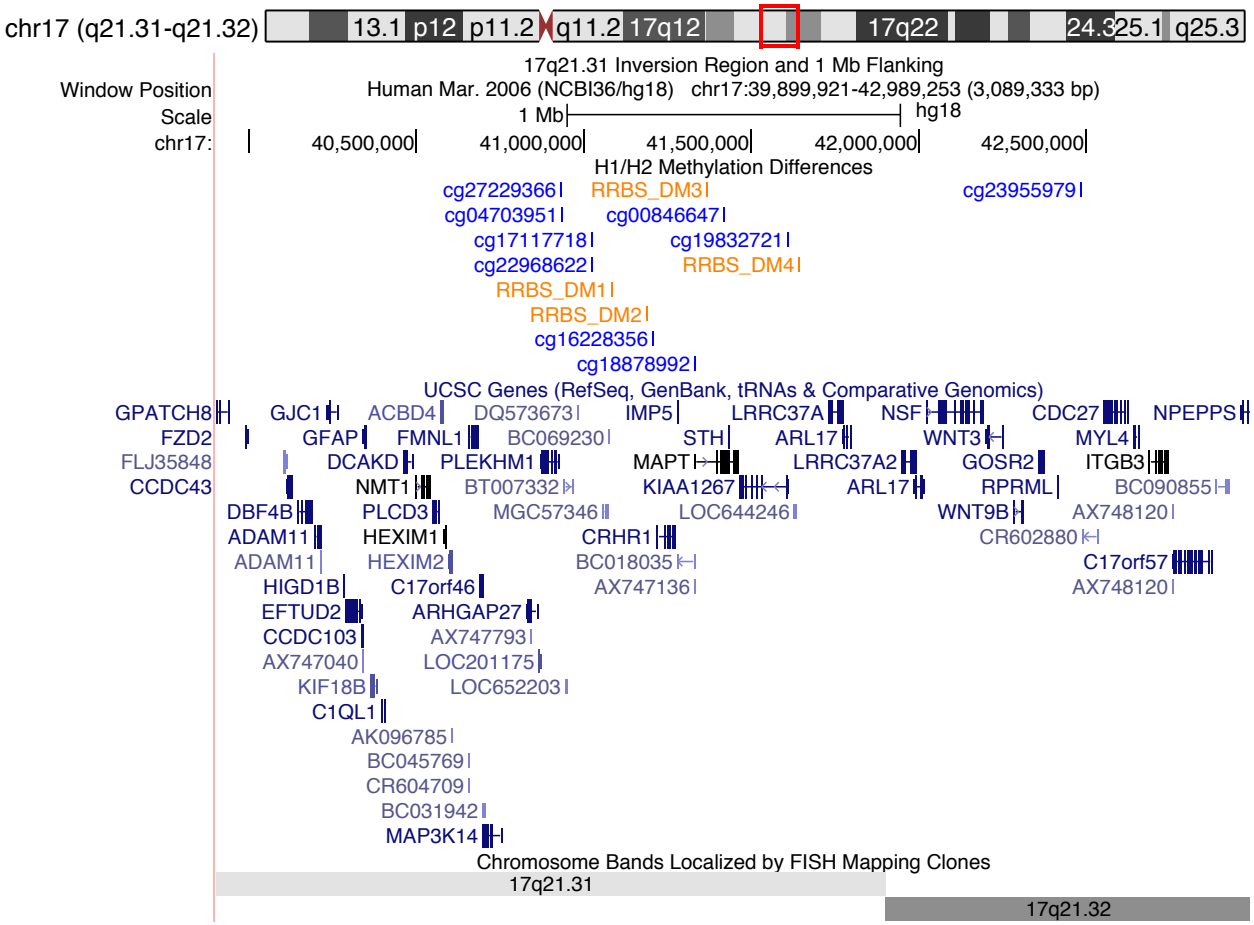
Supplementary Figure C-4: Methylation QTL analysis on the entire dataset (n=273), performed on 3 top DMPs identified when comparing H1 vs. H2 haplotypes. (a) scatterplot representing the R-squared for each of the SNPs in the 17q21.31 region associated with at least one of the three DMPs. Gray: not significant SNPs. (b) Manhattan plot representing p-values by chromosome. At each genomic location the smaller −log10 p-value from two datasets was plotted. A single cluster at 17q21.31 was identified for all three DMPs.



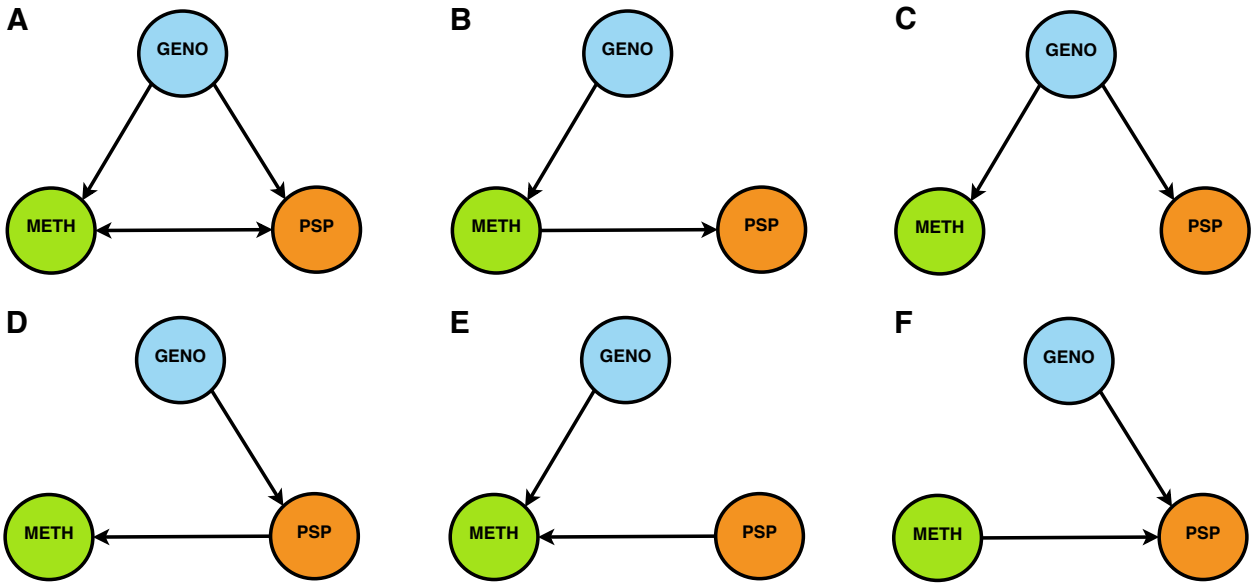
Supplementary Figure C-5: Correlation between average methylation fraction (β) values at common CpGs covered by both the Illumina HumanMethylation 450 k BeadChip Array and reduced representation bisulfite sequencing (RRBS), in seven samples from the study. The light blue points represent CpGs within the 17q21.31 cytoband.



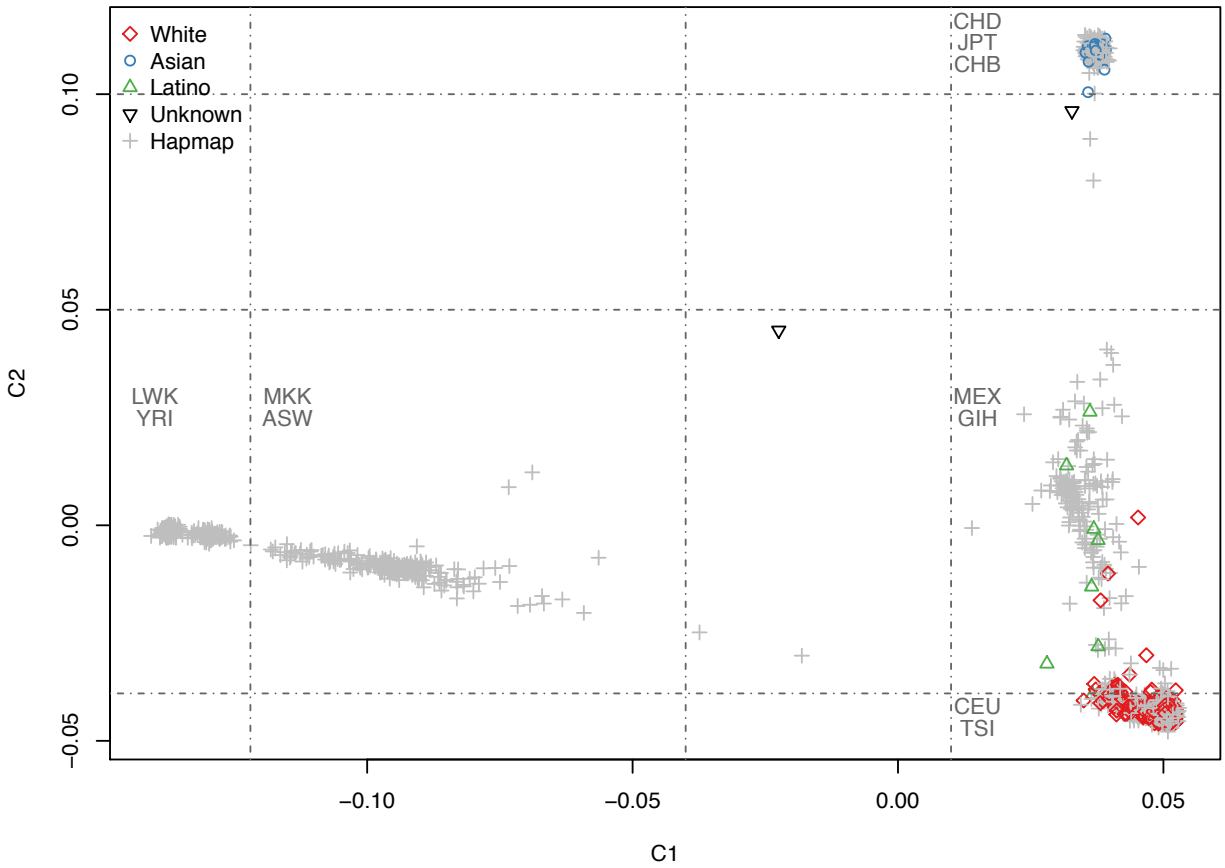
Supplementary Figure C-6: Correlation between average methylation fraction (β) values at common CpGs covered by both the Illumina HumanMethylation 450 k BeadChip Array and reduced representation bisulfite sequencing (RRBS), in two independent cohorts. The light blue points represent CpGs within the 17q21.31 cytoband.



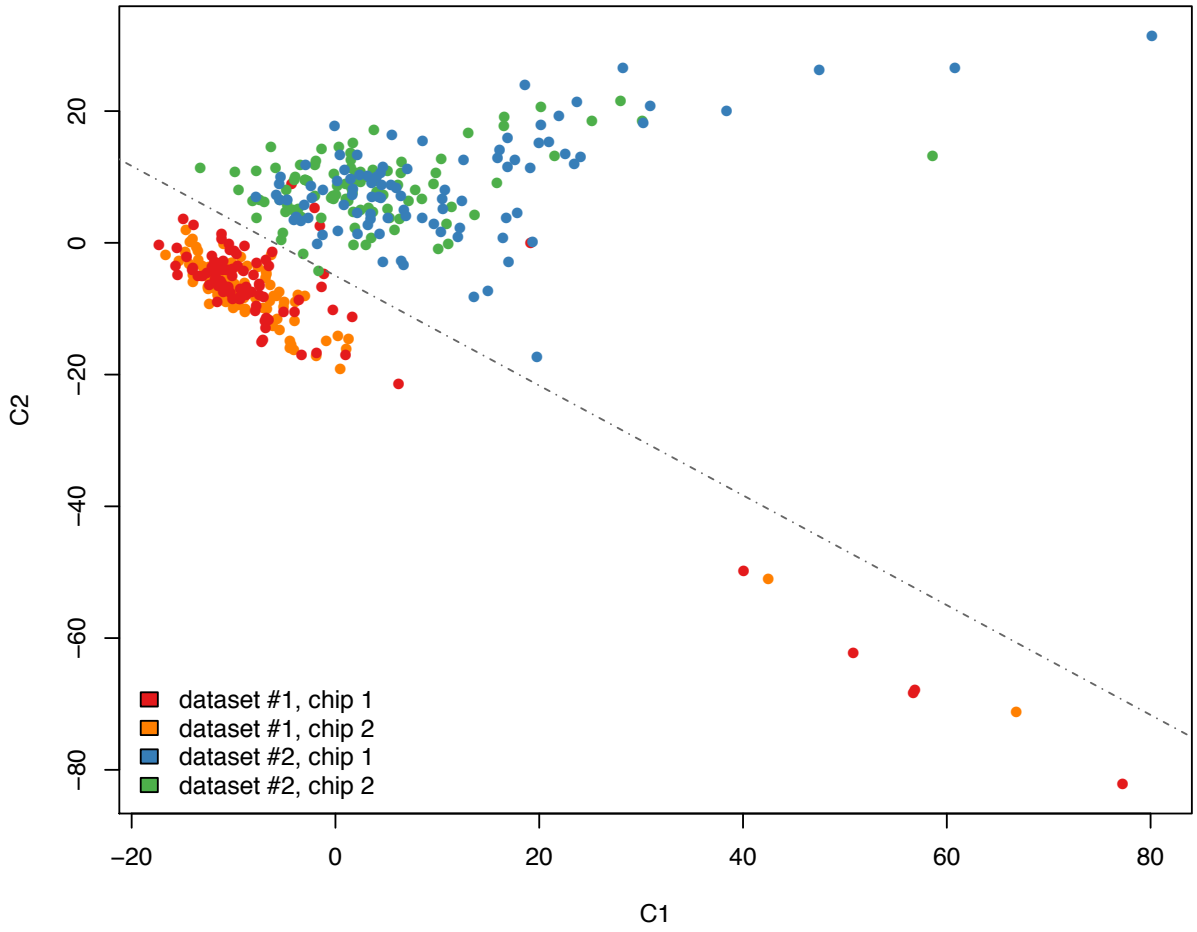
Supplementary Figure C-7: UCSC Genome Browser graphic and ideogram for the 17q21.31 inversion region, and 1 Mb of flanking sequencing on each side (which is also in linkage disequilibrium). Differentially methylated regions are depicted above the gene diagrams (blue: identified by Illumina HumanMethylation 450 k Array, and labeled with the Illumina Probe ID number; orange: identified by reduced representation bisulfite sequencing in an independent sample).



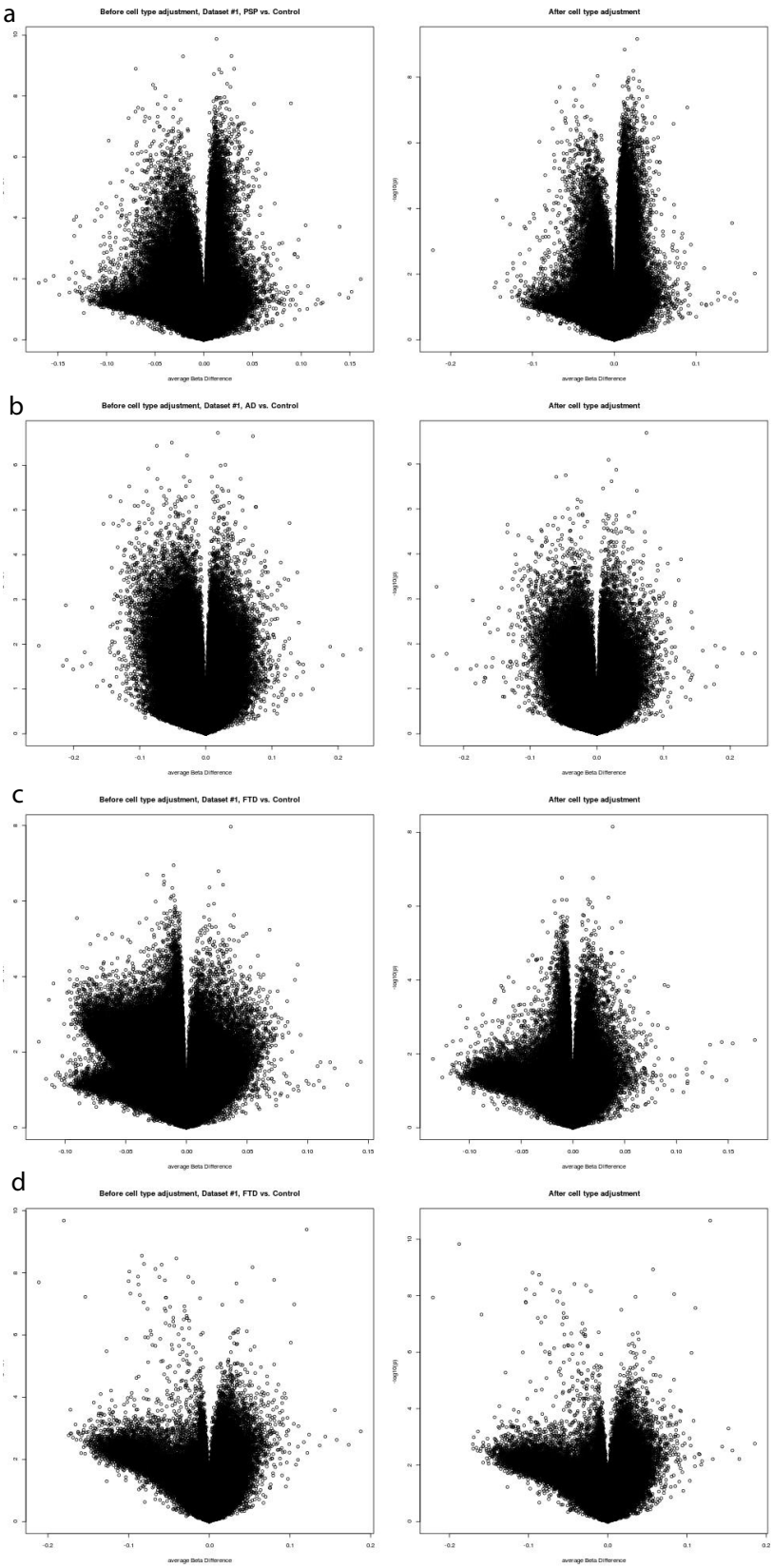
Supplementary Figure C-8: Causal models that explain the association between haplotype (HAPL), differentially methylated sites (METH), and PSP status (PSP). (a) Overview of the edges that are oriented by Network Edge Orienting (NEO) subroutine. The haplotype is anchored at the beginning of the causal diagram, as genotype precedes methylation and disease temporally (and thus, causally). NEO determines the most likely orientation of the remaining edges for each methylated region. (b) The "mediation model," in which the haplotype-associated effect is mediated by the intermediate step of methylation of a particular site. (c) An alternative model, in which the haplotype causes differential patterns of methylation independently from conferring disease risk. (d–f) The remaining three alternative causal models considered by NEO.



Supplementary Figure C-9: MDS plot representing the clustering of overlap between the SNP data in 273 samples from this study and Hapmap data. Samples are coded based on self-reported ethnicity.



Supplementary Figure C-10: Multidimensional scaling plot of Illumina 450 K methylation data showing a batch effect between two datasets. No obvious batch effect was observed within each dataset. SNP-containing probes, low-quality probes filtered out in each dataset, and sex chromosome probes were excluded from the analysis. The R function cmdscale was used for MDS analysis.



Supplementary Figure C-11: Volcano plots representing DMPs before and after cell type adjustment, in AD (a), PSP (b), FTD in dataset #1 (c), and FTD in dataset #2 (d).

Supplementary Table C-1: Demographic characteristics of the subjects enrolled in the study.

Dataset #1	Control	FTD	PSP
Number of subjects	93	55	40
Caucasian (%)	76.09 (n = 92)	79.63 (n = 54)	91.18 (n = 34)
Female (%)	57.78 (n=90)	45.83 (n=48)	29.17 (n=24)*
Age at onset ± SD	-	57.11±9.83 (n=46)	63.52 ± 7.00 (n=29)
Age at study ± SD	69.09±10.22 (n=93)	65.60±10.16 (n=55)*	70.36±7.44 (n=39)
APOE4 carriers (%)	20.65 (n=92)	25.45 (n=55)	11.76 (n=17)
APOE4 frequency (%)	11.96 (n=92)	13.64 (n=55)	5.88 (n=17)
17q21.31 H1 carriers (%)	95.65 (n=92)	94.44 (n=54)	100.00 (n=33)
17q21.31 H1 frequency (%)	79.89 (n=92)	77.78 (n=54)	96.97 (n=33)**
Dataset #2	Control	FTD	PSP
Number of subjects	92	73	3
Caucasian (%)	73.03 (n = 89)	80.82 (n = 73)	100 (n = 2)
Female (%)	68.66 (n=67)	50.00 (n=28)	0.00 (n=1)
Age at onset ± SD	-	60.49 ± 7.51 (n=41)	62.00 ± 1.41 (n=2)
Age at study ± SD	70.13±9.20 (n=92)	65.23±7.89 (n=73)***	65.33±5.03 (n=3)
APOE4 carriers (%)	15.38 (n=91)	24.66 (n=73)	
APOE4 frequency (%)	9.34 (n=91)	13.70 (n=73)	
17q21.31 H1 carriers (%)	91.30 (n=92)	97.18 (n=71)	100.00 (n=2)
17q21.31 H1 frequency (%)	80.98 (n=92)	86.62 (n=71)	100 (n=2)
Total	**Control**	**FTD**	**PSP**
Number of subjects	185	128	43
Caucasian (%)	74.59 (n = 181)	80.31 (n = 127)	91.67 (n = 36)*
Female (%)	62.42 (n=157)	47.37 (n=76)*	28.00 (n=25)**
Age at onset ± SD	-	58.70 ± 8.92 (n=87)	63.42±6.78 (n=31)

Age at study ± SD	69.61 ± 9.72(n=185)	65.39 ± 8.90 (n=128)***	70.00 ± 7.37 (n=42)
APOE4 carriers (%)	18.03 (n=183)	25.00 (n=128)	11.76 (n=17)
APOE4 frequency (%)	10.66 (n=183)	13.67 (n=128)	5.88 (n=17)
17q21.31 H1 carriers (%)	93.48 (n=184)	96.00 (n=125)	100.00 (n=35)
17q21.31 H1 frequency (%)	80.43 (n=184)	82.80 (n=125)	97.14 (n=35)**

Supplementary Table C-2: Breakdown of subjects by disease, and by H1/H2 genotype at the 17q21.31 locus.

Dataset #1	**AD**	**Control**	**FTD**	**PSP**
H1H1	-	59	33	31
H1H2	-	29	18	2
H2H2	-	4	3	0
Dataset #2	**AD**	**Control**	**FTD**	**PSP**
H1H1	9	65	54	2
H1H2	5	19	15	0
H2H2	1	8	2	0
Total	**AD**	**Control**	**FTD**	**PSP**
H1H1	9	124	87	33
H1H2	5	48	33	2
H2H2	1	12	5	0

Supplementary Table C-3: R-squared coefficients from multivariate linear regression model for 3 top PSP-related DMPs.

	cg23758822	cg22968622	cg12609785
Dx_Status	0.080**	0.009	0.012
H1_freq	0.081**	0.844***	0.602***
Ethnicity	0.003	0.001	0.009*
AGE	0.002	0.001	0.000

* p< 0.05, ** p < 0.01, *** p < 0.001

Supplementary Table C-4: Significant p-value computed using double-bootstrap standard error. 1000 bootstrap iterations were used for each of the two bootstrap methods of standard error estimation.

	Dataset #1				Dataset #2			
Cell type	H1 dominant	H1 recessive	FTD	PSP	H1 dominant	H1 recessive	FTD	AD
CD8T	0.295	0.462	0.043	0.097	0.047	0.001	0.071	0.014
CD4T	0.504	0.031	0.108	0.811	0.813	0.114	0.001	0.002
NK	0.824	0.096	0.014	0.065	0.019	0.002	0.004	0.537
B-cell	0.433	0.030	0.004	0.234	0.194	0.567	0.101	0.526
Mono	0.010	0.637	0.740	0.847	0.685	0.520	0.669	0.577
Gran	0.148	0.657	0.048	0.462	0.763	0.889	0.006	0.421

NK: natural killer, Gran: granulocytes, Mono: monocytes

Supplementary Table C-5: Breakdown of the 273 samples for which SNP array data and methylation data are available.

Dataset #1	**AD**	**Control**	**FTD**	**PSP**
H1H1	0	50 (39)	32 (27)	8+ 6 rs1052335 inferred (7)
H1H2	0	24 (21)	17 (16)	1 (1)
H2H2	0	3 (3)	3 (3)	0
Dataset #2	**AD**	**Control**	**FTD**	**PSP**
H1H1	9 (8)	38 (29)	41+ 2 rs1052553 inferred (34)	0
H1H2	4 (4)	12 (12)	13 (13)	0
H2H2	1 (1)	7 (6)	2 (2)	0
Total	**AD**	**Control**	**FTD**	**PSP**
H1H1	9 (8)	88 (68)	73 + 2 rs1052553 inferred (61)	8 + 6 rs1052335 inferred (7)
H1H2	4 (4)	36 (33)	30 (29)	1 (1)
H2H2	1 (1)	10 (9)	5 (5)	0

Supplementary Table C-6: Methylation QTL analysis for 3 DMPs within 17q21.31 in in 226 individuals of European descent.

Probe	Number of associated SNPs	R-squared (mean ± SD, range)	% within 17q21.31
cg22968622	113	0.773 ± 0.159, 0.255~0.982	99.12
cg17117718	113	0.727± 0.156, 0.264 ~ 0.8882	99.12
cg19832721	106	0.605± 0.087, 0.258 ~ 0.719	99.06

Supplementary Table C-7: Methylation QTL analysis for 3 DMPs within 17q21.31 in 273 individuals.

Probe	Number of associated SNPs	R-squared (mean ± SD, range)	% within 17q21.31
cg22968622	117	0.745 ± 0.22, 0.106~0.983	99.15
cg17117718	115	0.717± 0.198, 0.2 ~ 0.893	99.13
cg19832721	111	0.572± 0.134, 0.18 ~ 0.698	99.1

Supplementary Table C-8: Differentially methylated CpGs by genotype, found by reduced representation bisulfite sequencing.

Chromosome	Base	Position (hg18)	Context	p-value (uncorrected)	nearest gene
chr10	G	74087808	CG	5.15E-12	*MICU1* (intergenic)
chr17	C	41082844	CG	2.25E-12	*CRHR1* (intronic)
chr17	C	41182407	CG	4.56E-12	*CRHR1* (intronic)
chr17	C	41363555	CG	1.91E-12	*MAPT* (intronic)
chr17	G	41641866	CG	1.77E-13	*KIAA1267* (intronic)
chr17	C	52232615	CG	3.39E-09	*C17orf67* (intronic)
chr17	C	52232671	CG	7.71E-09	*C17orf67* (intronic)
chrX	C	107632332	CG	1.35E-11	*COL4A5* (intronic)
chrX	C	107632359	CG	6.72E-11	*COL4A5* (intronic)

Supplementary Table C-9: NEO predictions for each identified haplotype-dependent DMP.

Methylated Site	Within Gene	LEO.NB score (mediation model)†	Predicted Causal Model‡	Predicted Model P value	Model RMSEA
cg00846647	MAPT	-1.68	Meth<Hapl>PSP	0.019	0.143
cg04703951	-	-0.104	Meth<Hapl>PSP	0.089	0.0932
cg07870213	DND1	0.378	Hapl>Meth>PSP	0.578	0
cg16228356	-	-2.44	Meth<Hapl>PSP	0.000	0.234
cg17117718	-	0.85	Hapl>Meth>PSP	0.783	0
cg18878992	MAPT	-2.22	Meth<Hapl>PSP	0.001	0.211
cg19832721	KIAA1267	-0.874	Meth<Hapl>PSP	0.089	0.0931
cg22968622	-	1.51	Hapl>Meth>PSP	0.202	0.0537
cg23955979	-	-0.326	Meth<Hapl>PSP	0.081	0.0965

† LEO.NB score is defined as the base 10 logarithm of the ratio of p values for the mediation model and the next most likely causal model

‡ NEO chooses between five causal models; here, the arrows indicate the direction of causality.

Testing the 17q21.31 haplotype effect in patients and controls separately. We tested the dominant model (i.e. H1 carriers vs. H2/H2) in (1) controls and (2) FTD patients separately, and the recessive model (i.e. H1/H1 carriers vs. the remaining samples) in (3) AD patients, due to the small number of AD patients who were H2/H2 carriers. We were not able to perform a separate analysis for PSP subjects, as they were almost exclusively H1/H1 carriers.

H1 carriers were compared vs. non-carriers in control samples only (92 samples in each dataset, table S3). 110 DMPs were identified in dataset #1, 9 of which were located in the 17q21.31 region ($p = 2.605 \times 10^{-11}$, hypergeometric test). In dataset #2, 26 DMPs were identified, 10 of which were located within the 17q21.31 region ($p = 2.243 \times 10^{-20}$, hypergeometric test).

We then compared H1 carriers vs. non-carriers in FTD samples only. Sixteen DMPs were identified in dataset #1 (n=55), 5 of which were within the 17q21.31 region ($p = 2.697 \times 10^{-11}$, hypergeometric test). Five were identified in dataset #2 (n=73), 2 of which were in 17q21.31 region ($p = 5.883 \times 10^{-07}$, hypergeometric test).

Finally, due to the small number of AD patients who were H2H2 carriers, we tested the recessive model on AD samples in dataset #2 (n=15). The only DMP (cg22968622) was located within the 17q21.31 region.

In conclusion, in all cases we observed an overrepresentation of DMPs located within the 17q21.31 region, further supporting the notion that the observed changes in methylation are mostly due to genotype differences rather than disease status.

Impact of estimated relative cell counts in peripheral blood. The difference of cell type distribution between case and control were evaluated using the method developed by Houseman et al. and based on 500 loci whose methylation levels reflect the relative proportions of immune cells in unfractionated whole blood (Houseman et al., 2012; Koestler et al., 2013).

Briefly, we used methylation data available for 385 of the 500 loci (after quality control and probe removing). First, we estimated the blood cell type distribution for each sample using the methylation level of the 385 loci. Second, we applied a linear mixed-effect model considering (1) main blood cell types distribution as dependent variables; (2) disease status (or 17q21.31 haplotype), age, ethnicity, and gender as fixed effects; and (3) chip number as a random effect. We did not observe association of disease status or 17q21.31 haplotype with specific blood cell types (Table S5). Volcano plots (Figure S11) also showed no major differences in the number of DMPs when comparing before and after cell type adjustment.

Genome-wide methylation QTL analysis in the entire dataset (n=273). We performed a methylation QTL (methQTL) analysis in a subset of 273 individuals for whom whole-genome SNP data was available (Table S6). We assessed association of genetic variants with methylation levels at 3 CpGs within 17q21.31 (cg22968622, cg17117718, cg19832721) in each dataset. We identified on average 113 genome-wide significant signals (Bonferroni-adjusted $p \leq 0.05$), all located within the 17q21.31 region (Figure S4, Table S9). These variants accounted for a proportion of variability ranging between 10.6% and 98.3% (mean R-squared = 0.678, Figure S4, Table S10) further confirming that genetic variants at 17q21.31 are controlling methylation levels in *cis* in the same region.

Differential methylation analysis using the combined dataset (n=371 samples). After combining datasets #1 and #2, probes filtered in at least one dataset were removed, and the ComBat algorithm (Johnson et al., 2007) was applied to remove batch effects. 66,877 SNP-containing probes were excluded from further analysis, resulting in 397,528 probes analyzed in 371 samples.

We first compared H1 carriers (genotypes H1/H1 and H1/H2) vs. H2/H2 (dominant model) and identified 34 differentially methylated probes (BH-adjusted $p \leq 0.05$), 24 of which were located

within the 17q21.31 region ($p = 4.22 \times 10^{-53}$, hypergeometric test). Sixteen of these probes were also identified in both datasets when they were analyzed separately.

Second, we compared H1/H1 subjects vs. H2 carriers (genotypes H1/H2 and H2/H2, recessive model) and identified 45 differentially methylated probes (BH-adjusted $p \leq 0.05$), 36 of which were located within the 17q21.31 region ($p = 2.19 \times 10^{-81}$, hypergeometric test). Twenty-three of these probes were also identified in both datasets when they were analyzed separately. The 3 top DMPs (with an absolute aβD > = 0.1) identified in the two datasets when analyzed separately were also significantly differentially methylated in the combined analysis.

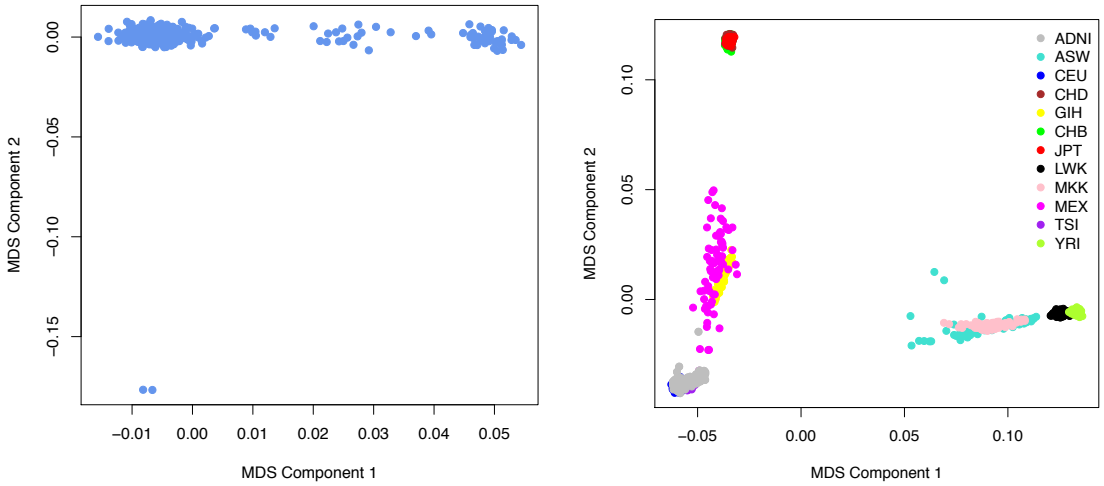
References

Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.

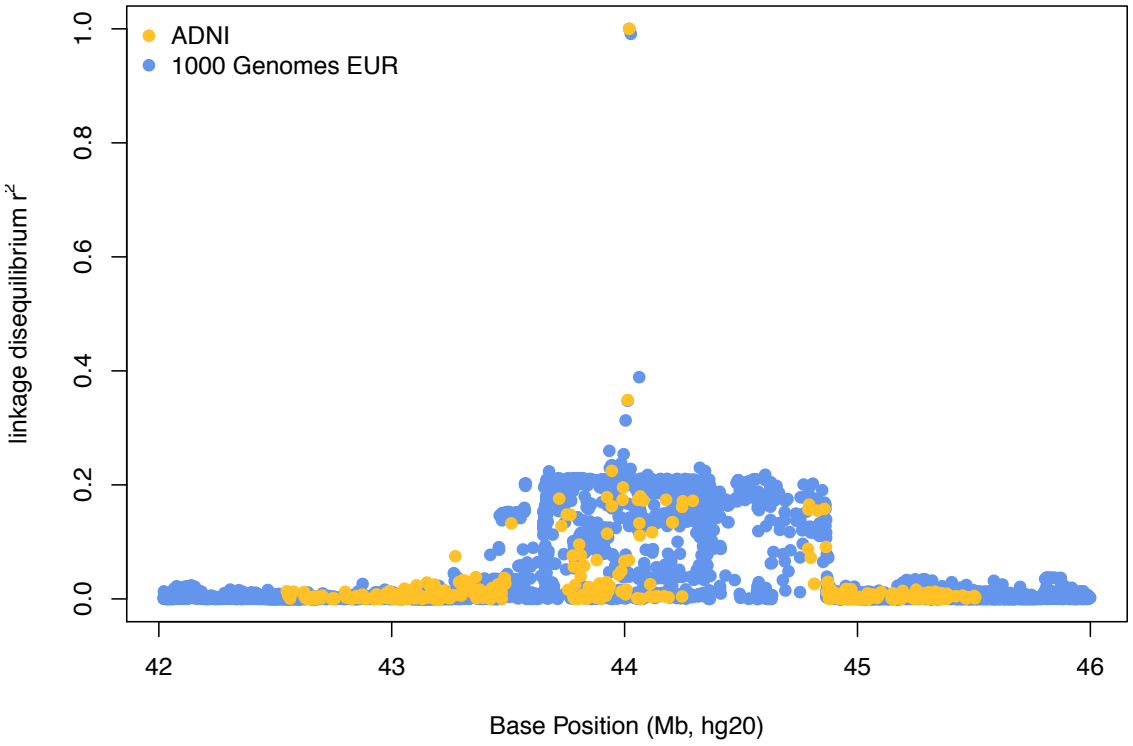
Koestler, D.C., Christensen, B.C., Karagas, M.R., Marsit, C.J., Langevin, S.M., Kelsey, K.T., Wiencke, J.K., and Houseman, E.A. (2013). Blood-based profiles of DNA methylation predict the underlying distribution of cell types. Epigenetics *8*, 816-826.

Appendix D: Supplementary Material for Chapter 6



a **b**

Supplementary Figure D-1: Cryptic relatedness and population substructure were checked with genomic identity-by-descent (IBD) and multidimensional scaling (MDS) components. (a) MDS plot of ADNI non-Hispanic Caucasian samples. Samples seemed to form loose clusters and two samples were outliers based on the second MDS component (at bottom of plot; 023_S_0058 and 023_S_0916), suggesting potential population substructure. To check for cryptic relatedness, which can confound GWAS studies, pairwise identity-by-descent fraction (π) between each pair of samples were calculated using PLINK. One related sample pair was identified (023_S_0058 and 023_S_0916, $\pi = 0.51$), which are probably first-degree relatives. No other cryptic relations were identified from the sample, at a threshold of $\pi > 0.2$. (b) MDS plot of ADNI samples overlaid on HapMap samples. The ancestry of the HapMap samples is shown by the point color. The outlying point represents Subject 116_S_1315 who is likely of mixed ancestry. Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; ASW, African ancestry in Southwest USA; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; CHB, Han Chinese individuals from Beijing, China; CHD, Chinese in Metropolitan Denver, Colorado; GIH, Gujarati Indians in Houston, Texas; JPT, Tokyo, Japan; LWK, Luhya in Webuye, Kenya; MEX, Mexican ancestry in Los Angeles, California; MKK, Maasai in Kinyawa, Kenya; TSI, Tuscans in Italy; YRI, Yoruba in Ibadan, Nigeria.



Supplementary Figure D-2: Linkage disequilibrium surrounding rs242557 in both the ADNI cohort and the 1000 Genomes cohort.

Supplementary Table D-1: SNPs associated with plasma tau at p-values between 10^{-5} and 10^{-6}.

CHR	SNP	MAF	Closest Gene	SNP Type/Location	Plasma Tau (*P* values)
9	rs7047280	0.406	*ELAVL2*	intergenic	1.20×10^{-5}
19	rs7251403	0.460	*CACNA1A*	intron	2.00×10^{-5}
5	rs7736147	0.360	*MAST4*	intron	2.11×10^{-5}
6	rs2874514	0.368	*PACRG*	intron	2.56×10^{-5}
16	rs12922111	0.224	*FOXL1*	intergenic	2.93×10^{-5}
10	rs2252372	0.454	*GOLGA2P6*	intron	3.19×10^{-5}
17	rs11867549	0.281	*MAPT*	intron	3.31×10^{-5}
9	rs4977611	0.394	*ELAVL2*	intergenic	3.32×10^{-5}
5	rs13175464	0.311	*MAST4*	intron	3.57×10^{-5}
18	rs319406	0.355	*ATP8B1*	intron	3.73×10^{-5}
12	rs7963143	0.392	*LMO3*	intergenic	4.22×10^{-5}
1	rs6660484	0.365	*TGFBR3*	intron	4.33×10^{-5}

Abbreviations: AD, Alzheimer's disease; *ATP8B1*, ATPase, Aminophospholipid Transporter, Class I, Type 8B, Member 1; *CACNA1A*, calcium channel, voltage-dependent, P/Q type, alpha 1A subunit; CHR, chromosome; ELAVL2, ELAV Like Neuron-Specific RNA Binding Protein 2; *FGD3*, FYVE, RhoGEF And PH Domain Containing 3; *FOXL1*, Forkhead Box L1; *GOLGA2P6*, Golgin A2 Pseudogene 6; *LMO3*, LIM Domain Only 3 (Rhombotin-Like 2); MAF, Minor allele frequency; *MAPT*, microtubule-associated protein tau; *MAST4*, Microtubule Associated Serine/Threonine Kinase Family Member 4; *PACRG*, PARK2 Co-Regulated; SNP, single nucleotide polymorphism; *TGFBR3*, Transforming Growth Factor, Beta Receptor III