Published in final edited form as:

Acta Neuropathol. 2017 May ; 133(5): 825–837. doi:10.1007/s00401-017-1693-y.

Shared genetic risk between corticobasal degeneration, progressive supranuclear palsy, and frontotemporal dementia

Jennifer S. Yokoyama[1,*,#], **Celeste M. Karch**[2,*], **Chun C. Fan**[3], **Luke W. Bonham**[1], **Naomi Kouri**[4], **Owen A. Ross**[4], **Rosa Rademakers**[4], **Jungsu Kim**[4], **Yunpeng Wang**[5], **Günter U. Höglinger**[6], **Ulrich Muller**[7], **Raffaele Ferrari**[8], **John Hardy**[8], **International FTD-Genomics Consortium (IFGC)**, **Parastoo Momeni**[9], **Leo P. Sugrue**[10], **Christopher P. Hess**[10], **A. James Barkovich**[10], **Adam L. Boxer**[1], **William W. Seeley**[1], **Gil D. Rabinovici**[1], **Howard J. Rosen**[1], **Bruce L. Miller**[1], **Nicholas J. Schmansky**[11], **Bruce Fischl**[11,12], **Bradley T. Hyman**[13], **Dennis W. Dickson**[4], **Gerard D. Schellenberg**[14], **Ole A. Andreassen**[6], **Anders M. Dale**[3,15], and **Rahul S. Desikan**[9,#]

[1]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA

[2]Department of Psychiatry, Washington University, St. Louis, MO, USA

[3]Department of Cognitive Sciences, University of California, San Diego, La Jolla, CA, USA

[4]Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, Florida 32224, USA

[5]NORMENT; Institute of Clinical Medicine, University of Oslo and Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

[6]Department of Neurology, Technical University of Munich, Munich, Germany and German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

[7]Institut for Humangenetik, Justus-Liebig-Universität, Giessen, Germany

[8]Department of Molecular Neuroscience, Institute of Neurology, UCL, London WC1N 3BG, UK

[9]Laboratory of Neurogenetics, Department of Internal Medicine, Texas Tech University Health Science Center, Lubbock, Texas, USA

[10]Neuroradiology Section, Department of Radiology and Biomedical Imaging, University of California, San Francisco

[11]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA

[12]Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, USA

[13]Department of Neurology, Massachusetts General Hospital, Charlestown, MA, USA

[#]Correspondence: Jennifer S. Yokoyama, PhD, Memory and Aging Center, University of California, San Francisco, 675 Nelson Rising Lane, Suite 190, San Francisco, CA 94158, jennifer.yokoyama@ucsf.edu, Phone: (415) 476-5565, Fax: (415) 502-7588. Rahul S. Desikan, MD, PhD, Neuroradiology Section, L-352, University of California, San Francisco, 505 Parnassus Avenue, San Francisco, CA, USA 92037-0841, rahul.desikan@ucsf.edu, Phone: (415) 353-1079, Fax: (415) 353-8593.

[*]equal contribution

[14]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[15]Departments of Radiology and Neurosciences, University of California, San Diego, La Jolla, CA, USA

Abstract

Corticobasal degeneration (CBD), progressive supranuclear palsy (PSP) and a subset of frontotemporal dementia (FTD) are neurodegenerative disorders characterized by tau inclusions in neurons and glia (tauopathies). Although clinical, pathological and genetic evidence suggests overlapping pathobiology between CBD, PSP, and FTD, the relationship between these disorders is still not well understood. Using summary statistics (odds ratios and p-values) from large genome-wide association studies (total n = 14,286 cases and controls) and recently established genetic methods, we investigated the genetic overlap between CBD and PSP and CBD and FTD. We found up to 800-fold enrichment of genetic risk in CBD across different levels of significance for PSP or FTD. In addition to NSF (tagging the *MAPT* H1 haplotype), we observed that SNPs in or near *MOBP*, *CXCR4*, *EGFR,* and *GLDC* showed significant genetic overlap between CBD and PSP, whereas only SNPs tagging the *MAPT* haplotype overlapped between CBD and FTD. The risk alleles of the shared SNPs were associated with expression changes in *cis*-genes. Evaluating transcriptome levels across adult human brains, we found a unique neuroanatomic gene expression signature for each of the five overlapping gene loci (omnibus ANOVA $p < 2.0 \times 10^{-16}$). Functionally, we found that these shared risk genes were associated with protein interaction and gene co-expression networks and showed enrichment for several neurodevelopmental pathways. Our findings suggest: i) novel genetic overlap between CBD and PSP beyond the *MAPT* locus; ii) strong ties between CBD and FTD through the *MAPT* clade, and; iii) unique combinations of overlapping genes that may, in part, influence selective regional or neuronal vulnerability observed in specific tauopathies.

INTRODUCTION

Collectively referred to as tauopathies, corticobasal degeneration (CBD), progressive supranuclear palsy (PSP) and a subset of frontotemporal dementia (FTD) spectrum disorders are neurodegenerative diseases defined by the intracellular, abnormal filamentous accumulations of the microtubule associated protein tau, encoded by *MAPT* [26]. Neuropathologically, tau deposits in CBD and PSP consist primarily of 4-repeat (4R) tau inclusions, and a large fraction of insoluble tau is present within neurons, astrocytes and oligodendrocytes [25]. Whereas CBD pathology predominantly affects forebrain structures (including the neocortex) and midbrain structures (including the substantia nigra), PSP pathology demonstrates a predilection for affecting hindbrain regions including the tectum, tegmentum, midbrain as well as the globus pallidus and diecephalon [14, 16]. In comparison, approximately half of FTD spectrum cases, encompassing behavioral variant FTD (bvFTD), semantic variant primary progressive aphasia (svPPA), non-fluent variant PPA (nfvPPA), and FTD overlapping with motor neuron disease (FTD-MND), are caused by underlying tau pathology. Approximately half of bvFTD clinical syndromes result from underlying Pick's disease, which is characterized by predominantly 3R tau inclusions (Pick

bodies) in the frontal, temporal and insular cortical gray matter [25]. The majority of nfvPPA clinical cases are due to underlying tau pathology, including PSP, CBD, and Pick's disease [9, 17, 21, 28]. Although neuropathologically distinct, there is considerable molecular and biomarker overlap between CBD, PSP and a subset of FTD [13]. Importantly, the biological basis of selective neuroanatomic or neuronal vulnerability within each tauopathy is poorly understood.

Genetic factors can offer novel insight into molecular mechanisms underlying disease risk. Recent genome-wide association studies (GWAS) have shown that single nucleotide polymorphisms (SNPs) within the H1 haplotype of the *MAPT* locus are associated with increased risk for CBD, PSP, and FTD [20, 24, 32]. Additionally, GWAS of pathologically confirmed CBD cases and from clinically and pathologically defined PSP cases have independently identified SNPs within *MOBP* as increasing disease risk [24] indicating that shared genetic risk may explain susceptibility to tau aggregates. Still, few, if any, studies have systematically evaluated genetic overlap between these three tauopathies.

Combining GWAS from multiple disorders and phenotypes provides insights into genetic pleiotropy (defined as a single gene or variant being associated with more than one distinct phenotype) and could identify shared genetic risk factors. In this study, using recently validated methods for systematically assessing genetic overlap, we investigated shared gene loci between tauopathies. Taking advantage of several large GWAS [15, 20, 24], we sought to identify SNPs *jointly* associated with CBD and PSP or FTD.

METHODS

Participant samples

We evaluated complete GWAS results in the form of summary statistics (p-values and odds ratios) for CBD, FTD and PSP (see Table 1). We assessed GWAS summary statistic data from 152 autopsy-proven CBD cases and 3,311 controls at 533,898 SNPs (Table 1). The CBD GWAS sample has been previously described in detail (see [24]). Briefly, the CBD cases were collected from eight institutions and controls were recruited from the Children's Hospital of Philadelphia Health Care Network. The National Institute of Health Office of Rare Diseases Research criteria were used for making a neuropathologic diagnosis of CBD [12]. We obtained phase I FTD-GWAS summary statistic data from the International FTD-Genomics Consortium (IFGC), which consisted of 2,154 clinical FTD cases and 4,308 controls with genotyped and imputed data at 6,026,384 SNPs (Table 1, for additional details, see [15]). The FTD dataset included multiple subtypes within the FTD spectrum: bvFTD, semantic dementia, progressive non-fluent aphasia, and FTD overlapping with motor neuron disease. We obtained publicly available PSP-GWAS summary statistic data from the NIA Genetics of Alzheimer's Disease Storage Site (NIAGADS), which consisted of 1,114 individuals with PSP (cases) and 3,247 controls (stage 1) at 531,451 SNPs (Table 1, for additional details see [20]). The relevant institutional review boards or ethics committees approved the research protocol of the individual GWAS used in the current analysis, and all human participants gave written informed consent.

Identification of shared risk loci -- conjunction FDR

Using recently developed statistical methods to evaluate shared genetic risk [1–3, 10, 11], we evaluated SNPs associating with CBD and FTD and CBD and PSP. For given associated phenotypes A and B, genetic 'enrichment' of phenotype A with phenotype B exists if the proportion of SNPs or genes associated with phenotype A increases as a function of increased association with phenotype B. To assess enrichment, we constructed fold-enrichment and conditional Q-Q plots of nominal $-\log_{10}(p)$ values for all CBD SNPs and for subsets of SNPs determined by the significance of their association with PSP and with FTD (for additional details see Supplemental Information). Enrichment seen can be directly interpreted in terms of true discovery rate (TDR = 1 – False Discovery Rate (FDR)) (for additional details see Supplemental Information and [1, 3, 10, 11]). Since 'genetic enrichment' is dependent on GWAS sample size [1], we applied our analysis focused on the CBD sample, the smallest available GWAS cohort.

To identify specific loci involved in both CBD and PSP or CBD and FTD, we computed conjunction FDR [1, 3, 36]. Conjunction FDR, denoted by $FDR_{trait1\&\ trait2}$ is defined as the posterior probability that a SNP is null for either phenotype or for both simultaneously, given the p-values for both traits are as small, or smaller, than the p-values for each trait individually (see [36] and Supplemental Information for more details). We used an overall FDR threshold of < 0.05. Additionally, we constructed Manhattan plots based on the ranking of conjunction FDR to illustrate the genomic location of the shared genetic risk loci.

Functional evaluation of shared risk loci

To assess whether the CBD, PSP, and FTD overlapping SNPs modify expression of genes, we evaluated *cis*-expression quantitative trait loci (eQTLs) in a publicly available dataset from neuropathologically confirmed normal control brains (UKBEC, http://braineac.org/) [29]. To minimize multiple comparisons, we focused on the average p-value derived from individual p-values from the cerebellum, frontal cortex, hippocampus, medulla, occipital cortex, putamen, substantia nigra, temporal cortex, thalamus, and white matter. To minimize risk of false positives, we used a Bonferroni corrected p-value (0.05/number of shared loci). We further assessed *cis*-eQTLs using a neurodegenerative dataset containing genotypic and gene expression data from Alzheimer's Disease [AD] (cerebellar expression GWAS (eGWAS) n = 197, temporal cortex eGWAS = 202) and PSP brains (cerebellar eGWAS = 98, temporal cortex eGWAS = 107) (for additional details see [38]). Finally, we also evaluated eQTLs using a blood-based dataset [33]. We used an analysis of covariance (ANCOVA) to test for association between genotypes and gene expression. We tested SNPs using an additive model.

Delineating differential regional expression of shared genetic risk variants

To evaluate whether the genes associated with the identified CBD, PSP, and FTD risk SNPs demonstrate a regionally specific pattern of expression across the human brain, we used the publicly available 6 human brain transcriptome data set from the Allen Brain Science Institute (for additional details see http://human.brain-map.org/ and [19]). Briefly, approximately 500 anatomically discrete samples were collected from cortex, subcortex, cerebellum, and brainstem of each brain and profiled for genome-wide gene expression

using a custom Agilent 8x60K cDNA array chip. For additional details regarding the dissection methods, quality control, and normalization measures on this dataset, please see [19] and http://human.brain-map.org/. For each gene associated with the shared SNPs, we downloaded expression values, calculated using z-score normalization. Using a repeated measures analysis of variance (ANOVA) (across the individual probes) as an omnibus test, we first investigated whether there was an overall effect of gene expression across all individual brain regions. Next, we calculated the mean of the expression values across each probe and across the six subjects and visualized these results using bar plots. Finally, for visualization, we mapped the mean expression values (across each probe and six subjects) onto the gray matter cortical surface using the Freesurfer image analysis package (www.freesurfer.net).

Gene ontology and network based functional association analyses

To identify significant enrichments in gene ontology associated with the PSP, CBD, and FTD pleiotropic genes, we used the *ToppGene* portal (https://toppgene.cchmc.org/). To evaluate potential protein and genetic interactions, co-expression, co-localization and protein domain similarity between the overlapping genes, we used GeneMANIA (www.genemania.org), an online web-portal for bioinformatic assessment of gene networks.

RESULTS

Shared genetic risk between tauopathies

We observed considerable SNP enrichment for CBD across different levels of significance with PSP and FTD (Figure 1). Using progressively stringent p-value thresholds for CBD SNPs (i.e. increasing values of nominal $-\log_{10}(p)$), we found up to 800-fold genetic enrichment using PSP and 400-fold genetic enrichment using FTD (Figure 1). Conditional Q-Q plots similarly demonstrated polygenic enrichment in CBD as a function of PSP and FTD (Supplemental Figure 1).

At a conjunction FDR < 0.05, we identified 5 SNPs that were associated with increased risk for CBD and PSP (Figure 2, Table 2, Supplemental Figures 2a–2e). These included: 1) rs199533 (exonic; *NSF*, CBD and PSP (and FTD), FDR $p = 3.85 \times 10^{-5}$); 2) rs1768208 (intronic; *MOBP*, CBD and PSP, FDR $p = 2.68 \times 10^{-3}$); 3) rs2011946 (intergenic; closest gene = *CXCR4*, CBD and PSP, FDR p = 0.038); 4) rs759162 (intronic; *EGFR,* CBD and PSP, FDR p = 0.045); and 5) rs7035933 (intronic; *GLDC*, CBD and PSP, FDR p = 0.045). We note that rs199533 is in strong linkage disequilibrium (LD) with SNP rs1800547, which tags the H1 haplotype of *MAPT* (pairwise $D' = 1$, $r^2 = 0.94$).

Although we excluded any *GRN* mutation carriers, given prior work illustrating TDP-43 pathology predominantly within svFTD, nfv-FTD and FTD-MND, one concern is that TDP-43, rather than tau pathology, may be influencing our results. To test this possibility, we performed sub-group analyses by excluding svFTD, nfv-FTD and FTD-MND individuals and specifically evaluated genetic overlap between CBD and bvFTD (FTD subtype which is enriched for tau pathology) and CBD and PSP. As illustrated in Supplemental Figures 3a–3b, limiting our analyses to the bvFTD subtype, we found similar

results as using the combined sporadic FTD cohort suggesting that our CBD-FTD findings may be strongly influenced by tau, rather than TDP-43, pathology.

We further assessed whether the observed genetic overlap between CBD, PSP and FTD was *polygenic* and generalizable across a number of loci or *non-polygenic* and driven by the *MAPT* region alone. Removing the chromosome 17 *MAPT*-region associated signal, consisting of all SNPs in r2 > 0.2 (based on 1000 Genomes Project LD structure) within 1 Mb of *MAPT* variants, we found considerable diminishment in genetic enrichment in CBD as a function of FTD (Supplemental Figures 4a–4b) further indicating that the observed overlap between CBD and FTD is non-polygenic and likely confined to the *MAPT* region. In contrast, evaluating genetic enrichment in CBD as a function of PSP after removing the *MAPT* region, we still found up to 80-fold enrichment suggesting that although the *MAPT* locus accounts for a large component of observed genetic overlap, variants beyond the *MAPT* region still contribute to shared genetic risk between CBD and PSP (Supplemental Figures 4a–4b).

eQTL analyses

Evaluating *cis*-eQTLs, at a previously established conservative Bonferroni corrected p-value $< 3.9 \times 10^{-5}$ [23], we found that rs199533 was significantly associated with *MAPT* ($p=2 \times 10^{-12}$) and a number of other transcripts in the *MAPT* region including *KIAA1267* ($p = 2.4 \times 10^{-24}$), *LOC1* ($p = 1.1 \times 10^{-23}$), *C17orf58* (p=2.1 × 10-6) and *LOC100294337* ($p=2.2 \times 10^{-6}$). As illustrated in Supplemental Figure 2a, rs199533 is in high LD with numerous SNPs within the *MAPT* haplotype block (including *MAPT*) that have a robust association with increased risk for tauopathy. Although rs199533 is linked to SNPs tagging the *MAPT* inversion region that distinguishes the H1 or H2 haplotype, reports vary as to the effects of *MAPT* haplotypes on gene expression (for a review of this topic see [23]). Whereas previous studies suggested that *MAPT* H1 is associated with increased production of pre-mRNA transcripts with exon 10 (making more 4R tau protein), more recent findings have indicated that the association with exon 10 maybe driven by a technical artifact [31]. Instead, it has been suggested that the *MAPT* H1 haplotype is associated with decreased expression of pre-mRNA transcripts containing exons 2/3 (2N tau protein). Using the UKBEC dataset, we assessed the effects of rs199533 on *MAPT* splicing by testing eQTLs in probes that specifically tag regions of the *MAPT* transcript, namely exon 3, exon 2, and exon 10. Applying a conservative Bonferroni corrected p-value $< 3.9 \times 10^{-5}$ [23] across the four probes, we found a specific and significant eQTL between rs199533 and exon 3 (average p-value across ten regions = 2×10^{-12}). In comparison, rs199533 and exon 2 (p-value = 0.02) and exon 10 (p-value = 0.04) demonstrated weaker eQTLs. Conditional analysis from the original PSP GWAS [20] found that rs242557 (in the promoter region of *MAPT*) was associated with increased disease risk, even after controlling for the H1/H2 clade, suggesting a *MAPT* sub-haplotype. We note that neither rs242557 nor a highly linked SNP, rs242562, are available in the UKBEC dataset (which we used to perform the exon-specific splicing analysis) and as such, we cannot exclude the possibility that a *MAPT* sub-haplotype is associated with increased inclusion of exon 10. Taken together, our results strongly suggest that variants within *MAPT* underlie the shared genetic risk observed on chromosome 17 and these variants alter expression of specific *MAPT* isoforms.

We also sought to investigate the functional impact of the other top SNPs for their role in gene expression of the loci they represent. As illustrated in Table 3, using the GTEX dataset, we found numerous significant eQTLs between rs2011946 and *MCM6* (within the *CXCR4* region, see Supplemental Figure 2c), rs1768208 and *MOBP* and rs7035933 and *GLDC*. Using the UKBEC dataset, we additionally evaluated the relationship between rs1768208 and *SLC5A38* expression (see Supplemental Figure 2b) and found a non-significant eQTL with the total probe (p-value = 0.56) further suggesting that our GWAS signal is localizing to the *MOBP* locus on chromosome 3. We also found that rs2011946 is associated with *CXCR4* expression in whole blood (p-value = 1.82×10^{-6}) [33]. This eQTL is consistent with the function of *CXCR4* in macrophages and microglia [27]. Although we cannot exclude other genes, our whole blood and GTEX eQTLs suggest that variants within the *CXCR4* region likely underlie our pleiotropic signal observed on chromosome 2. To further evaluate co-localization of eQTL and GWAS signals, we used COLOC (https://github.com/chr1swallace/coloc). Compatible with our eQTL results using the UKBEC and GTEX cohorts, our co-localization results confirm that the eQTL and GWAS signals are likely driven by the same causal SNPs (Supplemental Table 2).

Finally, we examined *cis*-eQTLs for rs199533 in a dataset containing genetic and expression data from control, AD and PSP human brain tissue [38]. We found numerous significant eQTLs in the cerebellum and frontal cortex (Supplemental Table 1). Both regions showed robust *MAPT* eQTLs, along with other loci (e.g., *LRRC37A*). We also found a significant eQTL between rs1768208 and *MOBP* within the temporal cortex (ILMN_1768947; p-value = 3.03×10^{-3}).

Delineating differential regional expression of shared variants

For each of the 5 shared risk genes, across the 26 brain regions, we found a significant overall effect for differential expression of *MAPT* (omnibus repeated measures ANOVA F (4,25) = 16.8, p-value = < 2.0×10^{-16}); *MOBP* (F (2,25) = 16.19, p-value = < 2.0×10^{-16}); *CXCR4* (F (2,25) = 16.19, p-value = < 2.0×10^{-16}); *EGFR* (F (4,25) = 14.6, p-value = < 2.0×10^{-16}); and *GLDC* (F (4,25) = 14.6, p-value = < 2.0×10^{-16}) indicating differential expression across brain regions. As illustrated in Figures 3 and 4, we found that each transcript showed a unique regional expression profile, all of which correspond to the neuroanatomical regions most affected in CBD, PSP, and FTD (Figure 5). Localized predominantly to the cerebral cortex, *MAPT* demonstrated the highest levels of expression within the temporal lobe, parietal lobe, parahippocampal gyrus, insula, and cingulate gyrus (Figures 3 and 4). Using the neocortical focused Braineac dataset, we independently confirmed this cortical pattern for *MAPT* regional expression (Supplemental Figure 5). *MOBP* showed the highest expression within the white matter and ventral thalamus (Figure 4). *CXCR4* demonstrated the highest expression within subcortical and deep gray structures (ventral thalamus, subthalamus, pontine tegmentum, and cerebellar nuclei) (Figure 4). *EGFR* showed the highest expression within the basal ganglia (globus pallidus, striatum), midbrain (mesencephalon and thalamus), and brainstem (myelencephalon and pontine tegmentum) (Figure 4). *GLDC* showed highest expression within the amygdala, cingulate gyrus, temporal lobe and claustrum (Figure 4). Interestingly, none of the 5 genes demonstrated

elevated expression within the hippocampal formation, a region preferentially affected in AD.

Gene ontology and gene-protein network analyses

At FDR < 0.05, we found that *MAPT, MOBP, CXCR4, GLDC,* and *EGFR* were enriched for multiple gene ontology (GO) molecular, cellular, and biological pathways (Table 4). Interestingly, we found that these shared loci were enriched for several neurodevelopmental process including telencephalic cell migration, forebrain cell migration, neuron projection morphogenesis, and neuron migration (q-value FDR < 0.05, Table 4). We found physical interaction and gene co-expression networks between *MAPT, MOBP, CXCR4, GLDC,* and *EGFR* (Figure 6). As illustrated in Supplemental Table 3, we found several 'network' based effects including physical interactions between *CXCR4* and *CXCL12* and *TLR2*, *GLDC* and *AMT,* and shared protein domains between *MAPT* and *MAP4*.

DISCUSSION

In this study, we identified genetic overlap between CBD, PSP, and FTD illustrating shared pathobiology between these diseases characterized primarily (PSP and CBD) or in part (FTD) by tau aggregation. Although *MAPT* and *MOBP* represented the strongest findings, we detected novel shared risk signals within *CXCR4*, *EGFR* and *GLDC*. Evaluating RNA levels across various brain regions, we detected a unique neuroanatomic gene expression signature for each of the five candidate genes. Functionally, we found that the genes representing these shared risk loci were associated with protein interaction and gene co-expression networks and showed enrichment for several neurodevelopmental pathways. Considered together, our results illustrate genetic overlap between tauopathies at and beyond the *MAPT* clade. These findings also suggest that a specific genetic combination of shared risk variants may determine selective vulnerability and thus influence whether a tauopathy-susceptible individual develops neurodegeneration within a frontotemporal (FTD), peri-rolandic (CBD), or midbrain/hindbrain (PSP) pattern.

By combining GWAS from multiple neurodegenerative diseases, our results suggest shared molecular pathways between these three tauopathies. Consistent with prior clinical and neuropathological observations that CBD and PSP exist along a 4R tauopathy spectrum [13], we found up to 800-fold genetic enrichment in CBD as a function of PSP illustrating considerable shared pathobiology that involves *MAPT* and non-*MAPT* regions. We also detected up to 400-fold genetic enrichment in CBD and FTD indicating overlap between 3R and 4R diseases, localized predominantly to the *MAPT* region. It should be noted that the GWAS for CBD and PSP utilized the same control series for analysis. Since we utilized summary statistics for the analysis, we cannot exclude the possibility that some of the shared genetic risk between these two disorders is driven by the controls.

By leveraging variants associated with increased risk for both PSP and FTD, we were able to identify variants associated with increased risk for CBD that have not been detected using a single phenotype approach. Beyond the known overlap between CBD and PSP at *MAPT* and *MOBP* [20, 24], we found novel genetic overlap within loci representing *CXCR4*, *EGFR* and *GLDC*. *MOBP* encodes myelin-associated oligodendrocyte basic protein, a Rab GTPase

critical for myelin sheath stabilization. Chemokine receptor *CXCR4* encodes a protein important in vascularization and cerebellar development. *EGFR* encodes epidermal growth factor receptor. *GLDC* encodes glycine dehydrogenase, a critical enzyme required for glycine degradation; mutations in *GLDC* cause glycine encephalopathy due to abnormally high levels of glycine in brain. Importantly, we found that the majority of these shared risk loci were either expressed in the brain (rs199533/*MAPT*, rs1768208/*MOBP*) or within macrophages/microglia (rs2011946/*CXCR4*). The biological roles of these different candidates coupled with their differential expression patterns in the brain support the idea that these genetic risk factors may promote specific neuroanatomical patterns of tauopathy when observed in different combinations. Considered together with prior work, our genetic findings together indicate that neuropathobiological alterations beyond the tau region are likely involved across tauopathies.

Evaluating RNA transcripts from six adult human brains, we identified a unique neuroanatomic gene expression signature for each of the five candidate genes. Building on prior studies evaluating *MAPT* expression within the normal adult brain [31], we found elevated total *MAPT* expression selectively within the neocortex. Across all evaluated regions, total *MAPT* RNA levels were highest within the temporal lobe, one of the earliest regions implicated in FTD ([4], see Figure 5). Interestingly, we did not find elevated total *MAPT* expression within the hippocampus, a brain region susceptible to tau pathology in middle and old age [7]. In contrast to *MAPT* and compatible with its known function in normal myelin formation [35], *MOBP* showed highest expression within white matter indicating that genetic variants associated with *MOBP* may play a role in selectively affecting white matter regions known to be involved with frontotemporal lobar degeneration [8], most strikingly in CBD. Interestingly, *CXCR4* and *EGFR* demonstrated elevated transcripts within the cerebellum and deep gray nuclei (thalamus and globus pallidus), regions more affected in CBD and PSP than FTD [6, 22, 34]. Considered together with our shared genetic risk findings, these gene expression results suggest that unique combinations of risk variants may influence selective regional vulnerability within the specific tauopathies.

Functionally, we found several gene-gene and gene-protein network-based associations between *MAPT, MOBP, CXCR4, GLDC,* and *EGFR*. Compatible with prior work showing the integral biophysical and molecular relationships between the chemokine CXCL12 (SDF-1) and its receptor CXCR4 [30], we detected a strong physical interaction between *CXCR4* and *CXCL12* in our bioinformatic analysis. Interestingly, although the CXCL12/CXCR4 axis has been described as a regulator of leukocyte trafficking and immune responses, recent evidence suggests that these chemokines are pleiotropic and may also be necessary for normal cell migration within the cerebral cortex [5] and cerebellum [18]. Building on these findings, we additionally found that our CBD, PSP, and FTD shared risk genes were enriched for several neurodevelopmental processes including neuron migration, tubulin complex, and neuron projection morphogenesis. Considered together, these results provide compelling support for shared genetic risk for tauopathies and suggest that genes critical for normal brain development may also influence brain degeneration.

The regional expression differences we observed occur in the context of a dataset derived from brain homogenates, which are composed of many different types of cells. While we are

unable to disentangle the exact contribution of each cell type to the observed regional expression differences, a separate dataset of transcriptomic information derived from single cell types from brain [37] provides indirect evidence that specific subtypes of cells may be driving these patterns in a gene-dependent manner (Fig S6). For example, *MAPT* expression is observed to be highest in neurons when compared to other cell types in the brain. In contrast, *MOBP* is far more highly expressed in oligodendrocytes versus any other cell type, consistent with its role in myelination. *CXCR4* is expressed most highly in microglia/macrophages as would be expected for a gene implicated in immune function. *EGFR* and *GLDC* both show highest expression in mature astrocytes, though the differences in expression in this cell type are not as marked as for some of the other, more specialized protein-encoding transcripts (i.e., *MOBP, CXCR4*). Together, the differential expression patterns of each gene candidate are consistent with their function and support a mechanism by which gene expression at both a regional and cell-specific level may contribute to selective vulnerability to particular cell types to tau pathology.

Our results indicate that common genetic variability at the *MAPT* haplotype and the *MOBP*, *CXCR4*, *EGFR* and *GLDC* genes may influence the regional pattern of volume loss—and thus disease phenotype—in individuals at risk for developing a specific tauopathy. For example, patients with *MAPT* and *CXCR4* susceptibility variants may be at greater risk for developing PSP than either FTD or CBD. Similarly, patients with *MAPT*, *MOBP* and *GLDC* may have a higher risk for developing CBD or FTD than PSP. Furthermore, the identification of a SNP that modifies a gene highly expressed in microglia (*CXCR4*) highlights the critical role of microglia in tau-mediated neurodegeneration.

In summary, beyond the extended *MAPT* haplotype, we found novel genetic overlap between CBD and PSP. Conversely, the extended *MAPT* haplotype represented the only pleiotropic locus for CBD and FTD. Each of these tauopathy-associated pleiotropic variants was associated with a unique neuroanatomic gene expression signature that may influence regional and neuronal selective vulnerability. Although no single common variant maybe informative clinically, the combination of these risk variants may in part drive whether a tauopathy-susceptible older individual develops CBD, PSP, or FTD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Primary support for data analyses was provided by the Larry L. Hillblom Foundation 2012-A-015-FEL and 2016-A-005-SUP (JSY), AFTD Susan Marcus Memorial Fund Clinical Research Grant (JSY), NIA K01 AG049152 (JSY), NIA K01 AG046374 (CMK), U24DA041123 (AMD, RSD), National Alzheimer's Coordinating Center (NACC) Junior Investigator (JI) Award (RSD), RSNA Resident/Fellow Grant (RSD), Foundation of ASNR Alzheimer's Imaging Grant (RSD), Alzheimer's Society Grant 284 (RF), and the Tau Consortium. GUH was supported by the Deutsche Forschungsgemeinschaft (DFG, HO2402/6-2 & Munich Cluster for Systems Neurology SyNergy), and the NOMIS Foundation (FTLD project). The PSP-GWAS was funded by a grant from the CurePSP Foundation, the Peebler PSP Research Foundation.

References

1. Andreassen OA, Djurovic S, Thompson WK, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. Am J Hum Genet. 2013; 92:197–209. DOI: 10.1016/j.ajhg.2013.01.001 [PubMed: 23375658]
2. Andreassen OA, McEvoy LK, Thompson WK, et al. Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes. Hypertension. 2014; 63:819–26. DOI: 10.1161/HYPERTENSIONAHA.113.02077 [PubMed: 24396023]
3. Andreassen OA, Thompson WK, Schork AJ, et al. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. PLoS Genet. 2013; 9:e1003455.doi: 10.1371/journal.pgen.1003455 [PubMed: 23637625]
4. Bang J, Spina S, Miller BL. Frontotemporal dementia. Lancet. 2015; 386:1672–1682. DOI: 10.1016/S0140-6736(15)00461-4 [PubMed: 26595641]
5. Borrell V, Marín O. Meninges control tangential migration of hem-derived Cajal-Retzius cells via CXCL12/CXCR4 signaling. Nat Neurosci. 2006; 9:1284–1293. DOI: 10.1038/nn1764 [PubMed: 16964252]
6. Boxer AL, Geschwind MD, Belfor N, et al. Patterns of brain atrophy that differentiate corticobasal degeneration syndrome from progressive supranuclear palsy. Arch Neurol. 2006; 63:81–6. DOI: 10.1001/archneur.63.1.81 [PubMed: 16401739]
7. Braak H, Del Tredici K. The preclinical phase of the pathological process underlying sporadic Alzheimer's disease. Brain. 2015; 138:2814–2833. DOI: 10.1093/brain/awv236 [PubMed: 26283673]
8. Chao LL, Schuff N, Clevenger EM, et al. Patterns of white matter atrophy in frontotemporal lobar degeneration. Arch Neurol. 2007; 64:1619–24. DOI: 10.1001/archneur.64.11.1619 [PubMed: 17998444]
9. Deramecourt V, Lebert F, Debachy B, et al. Prediction of pathology in primary progressive language and speech disorders. Neurology. 2010; 74:42–49. DOI: 10.1212/WNL.0b013e3181c7198e [PubMed: 19940270]
10. Desikan RS, Schork aJ, Wang Y, et al. Genetic overlap between Alzheimer's disease and Parkinson's disease at the MAPT locus. Mol Psychiatry. 2015; :1–8. DOI: 10.1038/mp.2015.6 [PubMed: 25648202]
11. Desikan RS, Schork AJ, Wang Y, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. Circulation. 2015; 131:2061–9. DOI: 10.1161/CIRCULATIONAHA.115.015489 [PubMed: 25862742]
12. Dickson DW, Bergeron C, Chin SS, et al. Office of Rare Diseases neuropathologic criteria for corticobasal degeneration. J Neuropathol Exp Neurol. 2002; 61:935–946. [PubMed: 12430710]
13. Dickson DW, Kouri N, Murray ME, Josephs KA. Neuropathology of frontotemporal lobar degeneration-Tau (FTLD-Tau). J Mol Neurosci. 2011:384–389.
14. Dickson DW, Rademakers R, Hutton ML. Progressive supranuclear palsy: pathology and genetics. Brain Pathol. 2007; 17:74–82. DOI: 10.1111/j.1750-3639.2007.00054.x [PubMed: 17493041]
15. Ferrari R, Hernandez DG, Nalls MA, et al. Frontotemporal dementia and its subtypes: A genome-wide association study. Lancet Neurol. 2014; 13:686–699. DOI: 10.1016/S1474-4422(14)70065-1 [PubMed: 24943344]
16. Forman MS, Zhukareva V, Bergeron C, et al. Signature tau neuropathology in gray and white matter of corticobasal degeneration. Am J Pathol. 2002; 160:2045–53. DOI: 10.1016/S0002-9440(10)61154-6 [PubMed: 12057909]
17. Grossman M, Xie SX, Libon DJ, et al. Longitudinal decline in autopsy-defined frontotemporal lobar degeneration. Neurology. 2008; 70:2036–2045. DOI: 10.1212/01.wnl.0000303816.25065.bc [PubMed: 18420483]
18. Haldipur P, Gillies GS, Janson OK, et al. Foxc1 dependent mesenchymal signalling drives embryonic cerebellar growth. Elife. 2014; doi: 10.7554/eLife.03962
19. Hawrylycz M, Miller JA, Menon V, et al. Canonical genetic signatures of the adult human brain. Nat Neurosci. 2015; 18:1832–1844. DOI: 10.1038/nn.4171 [PubMed: 26571460]

20. Höglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat Genet. 2011; 43:699–705. DOI: 10.1038/ng.859 [PubMed: 21685912]
21. Josephs KA, Duffy JR, Strand EA, et al. Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. Brain. 2006; 129:1385–1398. DOI: 10.1093/brain/awl078 [PubMed: 16613895]
22. Josephs KA, Whitwell JL, Dickson DW, et al. Voxel-based morphometry in autopsy proven PSP and CBD. Neurobiol Aging. 2008; 29:280–289. DOI: 10.1016/j.neurobiolaging.2006.09.019 [PubMed: 17097770]
23. Karch CM, Ezerskiy LA, Bertelsen S, Goate AM, (ADGC) ADGC. Alzheimer's Disease Risk Polymorphisms Regulate Gene Expression in the ZCWPW1 and the CELF1 Loci. PLoS One. 2016; 11:e0148717. [PubMed: 26919393]
24. Kouri N, Ross OA, Dombroski B, et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. Nat Commun. 2015; 6:7247.doi: 10.1038/ncomms8247 [PubMed: 26077951]
25. Kovacs GG. Invited review: Neuropathology of tauopathies: Principles and practice. Neuropathol Appl Neurobiol. 2015; 41:3–23. DOI: 10.1111/nan.12208 [PubMed: 25495175]
26. Lee VM, Goedert M, Trojanowski JQ. Neurodegenerative tauopathies. Annu Rev Neurosci. 2001; doi: 10.1146/annurev.neuro.24.1.1121
27. Lipfert J, Ödemis V, Wagner DC, Boltze J, Engele J. CXCR4 and CXCR7 form a functional receptor unit for SDF-1/CXCL12 in primary rodent microglia. Neuropathol Appl Neurobiol. 2013; 39:667–680. DOI: 10.1111/nan.12015 [PubMed: 23289420]
28. Mesulam MM, Weintraub S, Rogalski EJ, et al. Asymmetry and heterogeneity of Alzheimer's and frontotemporal pathology in primary progressive aphasia. Brain. 2014; 137:1176–1192. DOI: 10.1093/brain/awu024 [PubMed: 24574501]
29. Ramasamy A, Trabzuni D, Guelfi S, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci. 2014; 17:1418–28. DOI: 10.1038/nn.3801 [PubMed: 25174004]
30. Tiveron MC, Cremer H. CXCL12/CXCR4 signalling in neuronal cell migration. Curr Opin Neurobiol. 2008; 18:237–244. DOI: 10.1016/j.conb.2008.06.004 [PubMed: 18644448]
31. Trabzuni D, Wray S, Vandrovcova J, et al. MAPT expression and splicing is differentially regulated by brain region: Relation to genotype and implication for tauopathies. Hum Mol Genet. 2012; 21:4094–4103. DOI: 10.1093/hmg/dds238 [PubMed: 22723018]
32. Vandrovcova J, Anaya F, Kay V, et al. Disentangling the role of the tau gene locus in sporadic tauopathies. Curr Alzheimer Res. 2010; 7:726–734. CAR -100 [pii]. [PubMed: 20704554]
33. Westra H-J, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013; 45:1238–1243. [PubMed: 24013639]
34. Whitwell JL, Lowe VJ, Tosakulwong N, et al. [18 F]AV-1451 tau positron emission tomography in progressive supranuclear palsy. Mov Disord. 2016; doi: 10.1002/mds.26834
35. Yamamoto Y. Myelin-associated oligodendrocytic basic protein is essential for normal arrangement of the radial component in central nervous system myelin. Eur J Neurosci. 1999; 11:847–855. DOI: 10.1046/j.1460-9568.1999.00490.x [PubMed: 10103078]
36. Yokoyama JS, Wang Y, Schork AJ, et al. Association Between Genetic Traits for Immune-Mediated Diseases and Alzheimer Disease. JAMA Neurol. 2016; 94158:1–7. DOI: 10.1001/jamaneurol.2016.0150
37. Zhang Y, Chen K, Sloan SA, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci. 2014; 34:11929–11947. DOI: 10.1523/JNEUROSCI.1860-14.2014 [PubMed: 25186741]
38. Zou F, Chai H, Younkin C, et al. Brain Expression Genome-Wide Association Study (eGWAS) Identifies Human Disease-Associated Variants. PLoS Genet. 2012; 8:e1002707. [PubMed: 22685416]

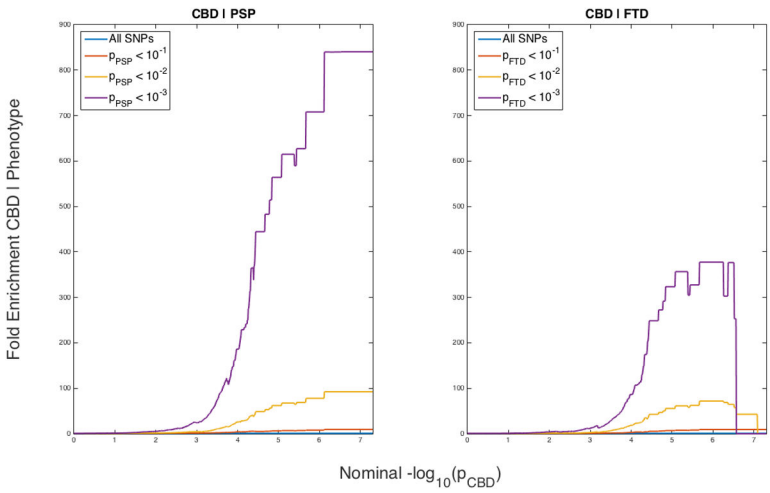

Figure 1.

Fold enrichment plots of enrichment versus nominal $-\log_{10}$ p-values (corrected for inflation) in corticobasal degeneration (CBD) below the standard GWAS threshold of $p < 5\times10^{-8}$ as a function of significance of association with progressive supranuclear palsy (PSP, left panel) and frontotemporal dementia (FTD, right panel) and at the level of $-\log_{10}(p) \geq 0$, $-\log_{10}(p) \geq 1$, $-\log_{10}(p) \geq 2$ corresponding to $p \leq 1$, $p \leq 0.1$, $p \leq 0.01$, respectively. Blue line indicates all SNPs.

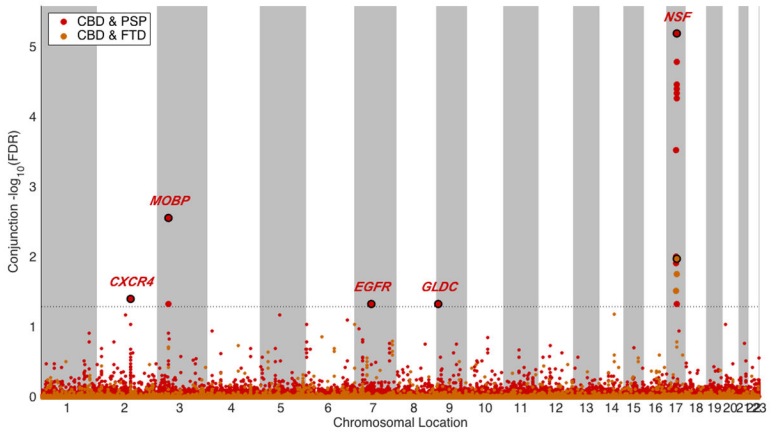

Figure 2.

'Conjunction' Manhattan plot of conjunction and conditional $-\log_{10}$ (FDR) values for corticobasal degeneration (CBD) (black) given progressive supranuclear palsy (PSP; CBD|PSP, red) and frontotemporal dementia (CBD; CBD|FTD, orange). SNPs with conditional and conjunction $-\log_{10}$ FDR > 1.3 (i.e. FDR < 0.05) are shown with large points. A black line around the large points indicates the most significant SNP in each LD block and this SNP was annotated with the closest gene, which is listed above the symbols in each locus.

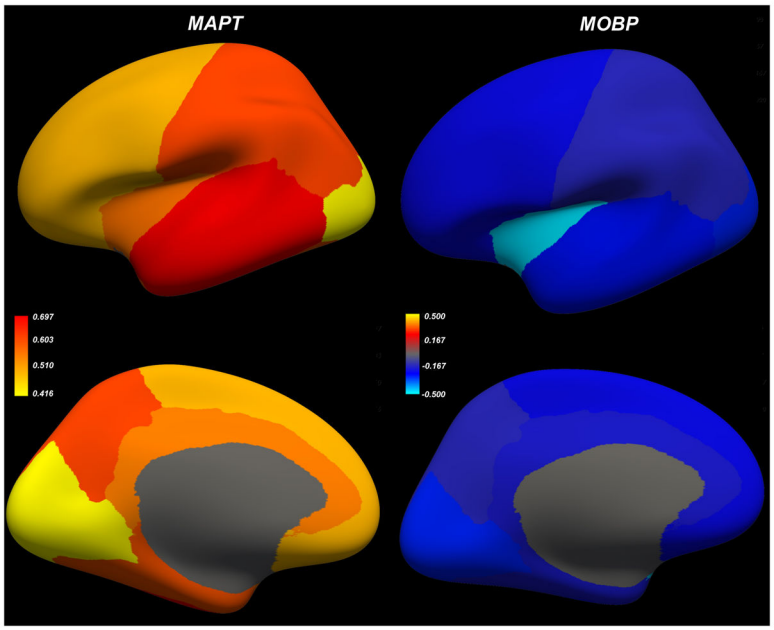

Figure 3.

Average regional gene expression of *MAPT* and *MOBP* from 6 postmortem subjects from the Allen Brain Science Institute mapped into a three dimensional reconstruction (‘inflated’ view) of the gray/white matter boundary of the cerebral cortex (fsaverage subject from FreeSurfer). Top panels illustrate the lateral view and bottom panels illustrate the medial view of the left cerebral hemisphere.

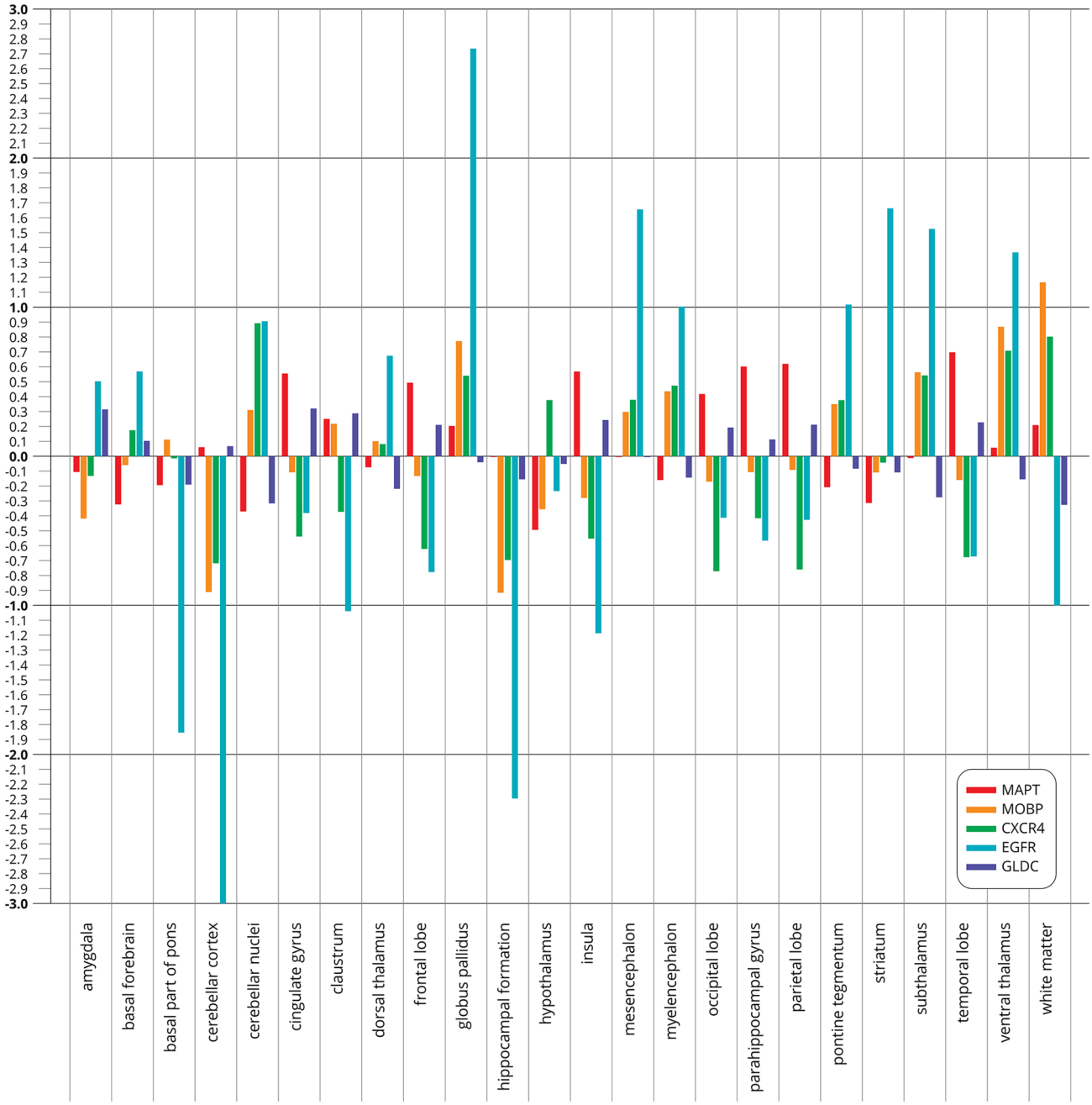

Figure 4.

Color-coded bar plots illustrating mean regional transcript levels for *MAPT*, *MOBP*, *CXCR4*, *EGFR* and *GLDC* using data from the Allen Brain Science Institute. Values represent z-scores averaged across the individual probes and across the 6 postmortem subjects.

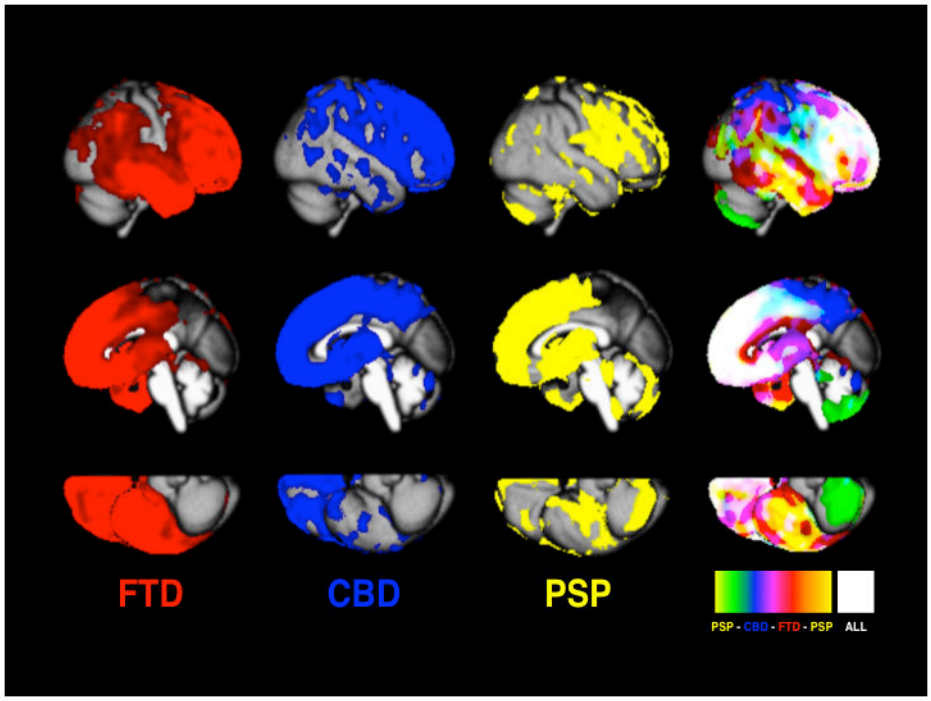

Figure 5.

Voxel based morphometry results illustrating atrophy patterns for FTD (red), CBD (blue), PSP (yellow), and overlap of all three tauopathies. Results are shown as T-map overlays ($p<0.001$) of volume loss in patients versus controls, and heat map reflects overlap between tauopathies. Atrophy patterns were derived from a cohort of individuals clinically diagnosed and pathologically confirmed to have CBD (n=6), FTD Pick's disease (n=5), or PSP (n=17), which was compared to a cohort of cognitively normal healthy older controls (n=180, mean ±SD 66.1±8.5 years). For detailed methods and results, see Supplemental Material.

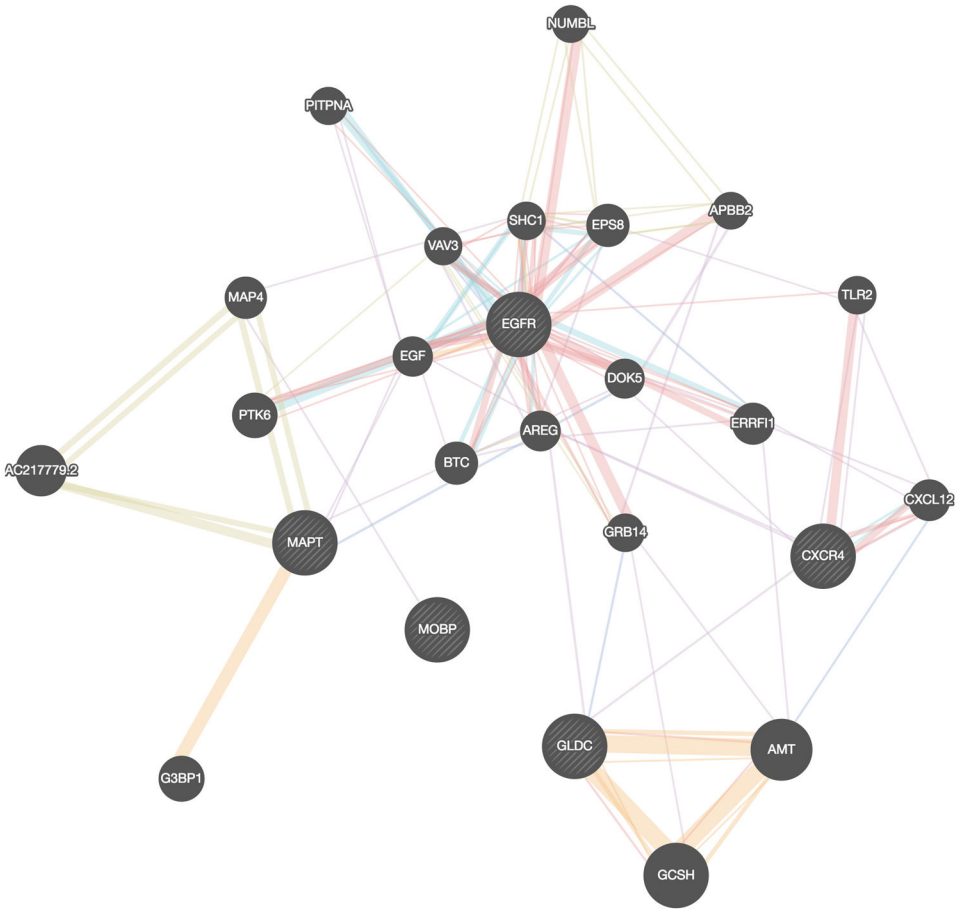

Figure 6.

Network interaction graph illustrating physical interactions (pink), co-expression (purple), predicted (orange), pathway (aqua), co-localization (blue), gene interactions (green) and shared protein domains (khaki) for *CXCR4, MOBP, EGFR, GLDC,* and *MAPT*.

Table 1

Summary data from all GWAS used in the current study

Disease/Trait	Total N	# SNPs	Reference
Corticobasal degeneration (CBD) – phase 1	3,463	533,898	Kouri N, et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. Nat Comm. 2015;6:7247.
ProgressiveSupranuclear Palsy (PSP) – phase 1	4,361	531,451	Höglinger GU, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat Genet. 2011:43:699–705.
Frontotemporal Dementia (FTD) – IFGC phase 1	6,462	6,026,384	Ferrari, R et al. Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neuro. 2014:13;686–99.

Table 2

Overlapping loci between CBD and PSP or FTD at a conjunction FDR < 0.05

SNP	Chr	Nearest Gene	Associated Phenotype	Associated Phenotype p-value	Min Conj FDR	CBD p-value
rs2011946	2	Intergenic; *CXCR4*	PSP	1.37E-04	3.84E-02	7.91E-04
rs1768208	3	Intronic; *MOBP*	PSP	4.90E-09	2.68E-03	1.04E-10
rs759162	7	Intronic; *EGFR*	PSP	8.20E-05	4.59E-02	1.04E-03
rs7035933	9	Intronic; *GLDC*	PSP	1.55E-04	4.59E-02	9.37E-04
rs199533	17	Exonic (K702K); *NSF,* tags *MAPT* H1 haplotype	PSP (and FTD)	5.56E-97	3.85E-5	3.00E-07

Acta Neuropathol. Author manuscript; available in PMC 2017 November 01.

Table 3

cis-eQTLs between tauopathy shared risk SNPs and associated genes across a variety of tissues from the UKBEC and GTEX cohorts.

SNP	Chr	Nearest Gene	eQTL			
			UKBEC		GTEX	
			P value	Gene	P value	Gene
rs2011946	2	*CXCR4*	3.20E-02#	*LCT*	3.22E-07	*MCM6* (Nerve Tibial)
rs1768208	3	*MOBP*	7.20E-01#	TTC21A	3.80E-29	*MOBP* (Nerve Tibial)
rs759162	7	*EGFR*	1.50E-01#	*MRPS17, ZNF713*	N/A	N/A
rs7035933	9	*GLDC*	4.10E-01#	*KIAA2026*	1.18E-11	*GLDC* (Nerve Tibial)
rs199533	17	*NSF*	1.70E-15	*KIAA1267, LOC100294337*	1.32E-18	*KANSL1-AS1* (Brain-Cortex)

*Note: Nerve Tibial where no CNS was available

#not significant

Table 4

Top functional associations of the shared risk genes. All associations are at q-value, FDR < 0.05.

GO: Biological pathway	GO: Cellular component	GO: Molecular function
Telencephalon Cell Migration (*EGFR, CXCR4*)	Multivesicular Body, Internal Vesicle Lumen (*EGFR*)	Glycine Dehydrogenase (Decarboxylating) Activity (*GLDC*)
Forebrain Cell Migration (*EGFR, CXCR4*)	Glycine Cleavage Complex (*GLDC*)	Oxidoreductase Activity (*GLDC*)
Neuron Projection Morphogenesis (*EGFR, CXCR4, MAPT*)	Shc-Egfr Complex (*EGFR*)	Enzyme Binding (*EGFR, CXCR4, MAPT, GLDC*)
Prolactin Secretion Regulation (*EGFR*)	Multivesicular Body, Internal Vesicle (*EGFR*)	Protein Phosphatase Binding (*EGFR, MAPT*)
Phospholipasea2 Activation (*EGFR*)	Growth Cone (*EGFR, MAPT*)	Epidermal Growth Factor-Activated Receptor Activity (*EGFR*)
Brain Development (*EGFR, CXCR4, MAPT*)	Site Of Polarized Growth (*EGFR, MAPT*)	Phosphatase Binding (*EGFR, MAPT*)
Neuron Migration (*CXCR4, MAPT)*	Tubulin Complex (*MAPT*)	Cytoskeletal Protein Binding (*EGFR, CXCR4, MAPT*)
Head Development (*EGFR, CXCR4, MAPT*)	Neurofibrillary Tangle (*MAPT*)	Epidermal Growth Factor Binding (*EGFR)*
Calcium-Mediated Signaling (*EGFR, CXCR4*)	Late Endosome (*EGFR, CXCR4*)	Structural Constituent Of Myelin Sheath (*MOBP*)
Glycine Catabolic Process (*GLDC*)	Early Endosome (*EGFR, CXCR4*)	Myosin Light Chain Binding (*CXCR4*)