RESEARCH ARTICLE

Calcium modulating ligand confers risk for Parkinson's disease and impacts lysosomes

Hanwen Zhang[1], Daniel Kargilis[1], Thomas Tropea[1] (ID), John Robinson[2], Junchao Shen[1], Eliza M. Brody[1], Ann Brinkmalm[3,4], Simon Sjödin[3], Adama J. Berndt[1], Marc Carceles-Cordon[1], EunRan Suh[2], Vivianna M. Van Deerlin[2], Kaj Blennow[3,4,5,6], Daniel Weintraub[7] (ID), Edward B. Lee[2], Henrik Zetterberg[3,4,8,9,10,11] & Alice S. Chen-Plotkin[1] (ID)

[1]Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

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

[3]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden

[4]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

[5]Paris Brain Institute, ICM, Pitie-Salpetriere Hospital, Sorbonne University, Paris, France

[6]Neurodegenerative Disorder Research Center, Division of Life Sciences and Medicine, and Department of Neurology, Institute on Aging and Brain Disorders, University of Science and Technology of China and First Affiliated Hospital of USTC, Hefei, China

[7]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[8]Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK

[9]UK Dementia Research Institute at UCL, London, UK

[10]Hong Kong Center for Neurodegenerative Diseases, Clear Water Bay, Hong Kong, China

[11]Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence

Alice S. Chen-Plotkin, Department of Neurology, Perelman School of Medicine, University of Pennsylvania, 150 Johnson Pavilion, Philadelphia 19104, PA, USA. Tel: 215-573-7193; Fax: 215-829-6606 E-mail: chenplot@pennmedicine.upenn.edu

Abstract

Objective: Several genetic loci known to confer risk for Parkinson's disease (PD) function in lysosomal pathways. We systematically screened common variants linked to PD risk by genome-wide association studies (GWAS) for impact on cerebrospinal fluid (CSF) proteins reflecting lysosomal function. **Methods**: Starting with 525 candidate gene-single nucleotide polymorphism (SNPs) pairs nominated by Mendelian randomization from published PD GWAS, we filtered SNPs for downstream evaluation, based on strength of association with PD and impact on brain gene expression. We genotyped top SNPs in 173 PD participants, adding three SNPs capturing variation at the *TMEM106B*, *CTSB*, and *RAB29* loci, encoding genes with known lysosomal function. In the same 173 individuals, we measured 15 CSF proteins (nine lysosomal proteins and six other proteins implicated in neurodegeneration) by parallel reaction monitoring mass spectrometry. We tested SNPs for association with lysosomal proteins. For our top SNP associating with multiple lysosomal proteins, we characterized expression of its target gene *CAMLG* in human brain tissue. **Results**: Sixteen SNPs emerged from our analysis of GWAS-nominated loci. Genotypes at rs12657663 (*CAMLG*) associated with CSF levels of multiple lysosomal markers (cathepsin F, cathepsin L, hexosaminidase B, and tripeptidyl peptidase I) and genotypes at rs7910668 (*ITGA8*) with CSF levels of cathepsin B. The protein encoded by *CAMLG*, calcium modulating ligand (CAML), is highly expressed in neurons of multiple human brain regions, with higher expression in Lewy body disease cases. **Interpretation**: Systematic analysis of PD risk loci nominates *CAMLG* as a neuronally expressed risk gene with effects on lysosomes.

Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder marked by both motor symptoms, including bradykinesia, resting tremor, and rigidity; and nonmotor symptoms, such as cognitive impairment.[1] The second most common neurodegenerative disease, PD affects more than 5 million people worldwide, including 1% of the population over 60 years, with further increases projected in the future.[2,3] Hence, an urgent, unmet need is to identify specific targets for disease modification.

Lysosomal dysfunction has been implicated in the pathogenesis of PD.[4,5] Variants in *GBA1* impact 10% of PD individuals.[6–8] Many PD-associated *GBA1* variants decrease function of the encoded protein glucocerebrosidase, a lysosomal hydrolase; this lysosomal impact is believed to increase accumulation of misfolded alpha-synuclein (aSyn), the hallmark protein that characterizes the Lewy body pathology found in PD.[9,10] Variants in *LRRK2* impact ~2% of PD individuals, and emerging studies link the encoded kinase to lysosomal function as well.[11] In addition to variants in *GBA1* and *LRRK2*, both autosomal dominant causes of PD with incomplete penetrance, several other genetic loci linked to PD or endophenotypes in PD also have known effects on lysosomes. Single nucleotide polymorphisms (SNPs) in *TMEM106B*, linked to rate of cognitive decline in PD, impact *TMEM106B* expression levels[12–15]; *TMEM106B* over-expression leads to defects in lysosomal acidification and degradative capacity, while C-terminal fragments of the encoded protein (transmembrane protein 106B, or TMEM106B) form amyloid fibrils that accumulate in the lysosomes of neurodegenerative disease brain tissue.[16,17] Variants in *CTSB*, encoding the lysosomal enzyme cathepsin B, have been linked with PD risk by genome-wide association studies (GWAS) and also act as modifiers of age at onset in *GBA1*-associated PD.[18,19] Genetic studies have also linked variation at *RAB29* with PD risk, and the encoded protein functions in endolysosomal sorting.[20,21] These genetic data suggest that lysosomal dysfunction may be broadly involved in PD pathogenesis, and PD risk genes acting in lysosomal pathways may be promising therapeutic targets.

Here, we systematically interrogate PD GWAS loci to link risk loci to specific target genes that may impact lysosomal function in human PD. To do this, we filtered 525 SNP-target gene pairs derived from a Mendelian randomization (MR) analysis of PD GWAS data, retaining candidates with strong associations with PD and sizeable impact on target gene expression[22] (Fig. 1). To the top 16 SNPs obtained in this manner, we added SNPs at the *TMEM106B*, *CTSB*, and *RAB29* loci, based on known effects on lysosomes for these target genes. We then tested genotype at these 19 loci for impact on lysosomal proteins in the cerebrospinal fluid (CSF) of 173 PD individuals, yielding two lead SNPs. For the top candidate, rs12657663, a missense SNP in *CAMLG*, we performed an initial characterization in human brain tissue from PD and control subjects.

Materials and Methods

Participants

Individuals diagnosed with PD based on the UK Brain Bank criteria[23] aged 50 and older and neurologically normal controls were enrolled at the University of Pennsylvania Parkinson's and Movement Disorders Center (Philadelphia, PA) as part of the University of Pennsylvania (UPenn) Udall-U19 Cohort.[24] This study was approved by the University of Pennsylvania Institutional Review Board. Informed consent was obtained at study enrollment. Lumbar puncture was performed at study enrollment, and the collected CSF samples were stored at −80°C in 0.5 mL per aliquot. DNA was isolated from whole blood or saliva samples from all participants, as previously described, and the data were stored in the Penn Integrated Neurodegenerative Disease Database.[25]

CSF analysis by PRM-MS

Forty-five individual peptides, covering 15 proteins of interest (one to four peptide fragments per protein), were measured by parallel reaction monitoring mass spectrometry, optimized from previously published methods.[26] CSF samples were dissolved in 50 mM NH_4HCO_3 before separation by Dionex UltiMate 3000 standard-LC system (Thermo Fisher Scientific Inc., Waltham, MA, USA). Single microscans were acquired in PRM mode with higher-energy collisional dissociation. Measurements of endogenous peptides were determined based on ratio to a corresponding heavy stable-isotope labeled standard.

For each protein of interest, total peptide measures were calculated across all measured individual peptide fragments. To maintain the relative distance between measurements across samples while also giving each peptide equivalent weight in the "combined" value, individual fragment concentrations were first normalized against the average total peptide concentrations measured in all samples. Next, the normalized peptide fragment concentrations from each protein were averaged to produce one total value for each protein that was used in our analysis.

Genotyping

Genotyping at candidate single nucleotide polymorphisms (SNP) was performed via NeuroX genotyping platform
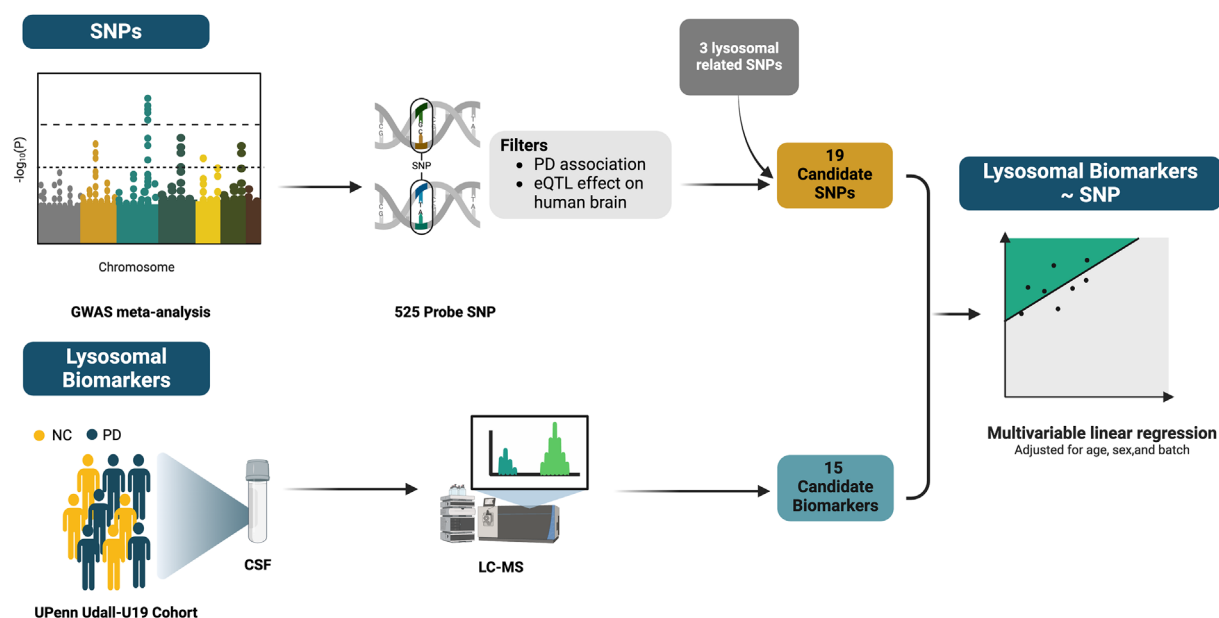
Figure 1. Experimental overview. SNPs nominated in a previously published Mendelian randomization analysis of PD genome-wide association study risk loci were filtered by their PD association ($P < 1e\text{-}08$) and expression quantitative trait locus (eQTL) effect on human brain ($\beta > 30\%$) to yield 19 top candidate SNPs. Fifteen candidate proteins were measured by mass spectrometry in CSF samples from PD and neurologically normal control (NC) individuals. SNPs were then assessed for their impact on lysosomal protein levels by multivariable linear regression.

and GSA (Illumina Infinium Global Screening) chip screening arrays.[25] For GWAS SNPs that were not directly genotyped on these arrays, proxy SNPs were selected ($D' > 0.9$, with R^2 as close to 1 as possible).[27] See Table 1 for GWAS SNP and proxy SNP identifiers, D', and R^2. Subjects were screened for *GBA1* variants as previously described,[28] using long-range polymerase chain reaction amplification, followed by sequencing of all exons of the *GBA1* gene.

Immunohistochemistry

Formalin-fixed, paraffin-embedded hippocampus, frontal cortex, temporal cortex, thalamus, midbrain, and cerebellum samples were obtained from the brain bank of the UPenn Center for Neurodegenerative Disease Research (CNDR). Case characteristics are reported in Table S1. 6 μm sections were deparaffinized and rehydrated in xylenes and a descending 100%, 95%, 80%, and 70% EtOH series. Endogenous peroxidases were quenched in 70% MeOH and 30% H_2O_2 solution for 30 min. Slides were 95°C water-bath heated in citric acid Antigen Unmasking Solution (Vector Laboratories) for 30 min. After cooling, slides were rinsed in Tris (0.1 M pH 7.6) and blocked (Tris/2%FBS/3%BSA) for 5 min. Sections were incubated overnight at 4°C in the primary CAMLG antibody (NovusBio NBP2-33789; 1.07 μg/mL). Once washed with Tris and blocked for

5 min, sections were incubated for 1 h at room temperature in the biotinylated rabbit secondary antibody (Vector Laboratories, 1:1000). VECTASTAIN ABC Standard (Vector Laboratories) was applied for 1 h at room temperature followed by ImmPACT DAB (Vector Laboratories) for 45 s. Sections were counterstained with Harris Hematoxylin (Thermo Scientific) for 50 s. Slides were dehydrated in an ascending 70%, 80%, 95%, and 100% EtOH series and then xylenes. Coverslips were sealed with Cytoseal (Thermo Scientific). Images were acquired on an Olympus VS200 slide scanner. An operator blinded to case status selected slide sections with comparable levels of the hippocampal formation containing the dentate gyrus, and Ammon's horn, then annotated a region of interest containing these hippocampal structures. HALO software (Indica labs, v3.5.3577.214, area quantification package v2.4.2) was used to identify pixels with positive staining vs. pixels with negative staining in the selected area of interest using the Area Quantification module, and the percentage of positively stained area was calculated. See Fig. S5 for details.

Statistical analysis

All statistical analyses and figure creation were conducted in R (version) (http://www.r-project.org); R-scripts are available in the Supplement and GitHub (https://github.

Table 1. SNP information summary.

Sentinel SNP	eQTL gene	Assay	Proxy SNP	D′	R^2	Coordinates of measured SNP	Alleles	MAF of measured SNP
rs12657663	CAMLG	NeuroX		1	1	chr5:134741122	(G/A)	0.0965
rs7910668	ITGA8	NeuroX		1	1	chr10:15506926	(G/A)	0.1324
rs858274	GPNMB	NeuroX		1	1	chr7: 23255049	(T/C)	0.4072
rs4631423	CTSB	NeuroX		1	1	chr8: 11841770	(C/T)	0.1765
rs2305880	STX1B	NeuroX		1	1	chr16: 30988141	(T/C)	0.358
rs11248057	GAK	GSA	rs6599388	0.9493	0.8182	chr4: 945299	(C/T)	0.2879
rs71513892	BIN3	GSA	rs7843828	1	0.9763	chr8: 22606110	(C/T)	0.3327
rs1876543	FAM47E	GSA	rs4859655	1	0.9797	chr4: 76273511	(T/G)	0.3901
rs9271548	HCG23	GSA	rs707951	0.9068	0.7365	chr6: 32642036	(T/C)	0.3293
rs11645306	LAT	NeuroX	rs7140	1	1	chr16: 28984436	(C/A)	0.2918
rs77823616	MMRN1	GSA	rs76572204	1	0.7696	chr4: 89878300	(G/A)	0.0092
rs60492656	TMEM175	NeuroX		1	1	chr4: 881220	(C/T)	0.1803
rs10756899	SH3GL2	NeuroX	rs10810829	1	0.8357	chr9: 17697097	(G/T)	0.2537
rs17191234	HLA-DQA2	NeuroX	rs614348	1	0.6912	chr6: 32606100	(T/C)	0.3357
rs7653459	ARIH2	GSA	rs7647812	1	1	chr3: 48795914	(G/A)	0.3482
rs3772034	KCNIP3/PROM2	GSA	rs4555374	0.9495	0.8582	chr2: 94973759	(G/A)	0.2338
rs1990622	TMEM106B	GSA	rs6966915	1	1	chr7: 12226362	(C/T)	0.4107
rs199502	MAPT	GSA	rs199515	0.9672	0.9058	chr17: 46779275	(G/C)	0.1857
rs7522056	RAB29	NeuroX	rs1935025	1	0.7392	chr1: 205722956	(C/A)	0.2463

Nineteen PD risk SNPs were investigated for their effect on lysosomal protein levels. For each SNP, the target gene indicated by eQTL analyses is shown, along with the assay used for genotyping. For 13/19 SNPs, proxy SNPs (that were directly genotyped on the assay listed) were used, rather than the sentinel SNP identified by genome-wide association. For these, the D′ and R^2 values between sentinel and proxy SNPs are shown. Coordinates, alleles, and the minor allele frequency (MAF) for the genotyped SNP (Proxy or Sentinel) are shown. Coordinates are based on GRCh38 (Genome Research Consortium human build 38), and the MAF is based on European populations in the database of Genotypes and Phenotypes (dbGaP).

com/ChenPlotkinLab/CAMLG). The mean, median, and standard deviations were calculated for age of individuals and age of onset across PD cognition categories: PD-cognitively normal (PD-CN), PD-mild cognitive impairment (PD-MCI), and PD-dementia (PDD). The average total DRS scores across PD cognition categories were also calculated.

Multivariable linear regression models were used to assess associations between genotype at each SNP (coded additively by number of major alleles, 2 = major allele homozygosity, 1 = major allele heterozygosity, 0 = minor allele homozygosity) and the CSF concentrations of the "combined" peptide fragments of each protein measured. All models were adjusted for covariates including age at sample collection, sex, and plate batch: protein levels were regressed on the SNP, age, sex, and plate variables. The Benjamini–Hochberg (BH) method was applied to analyses to account for multiple comparisons, and adjusted alpha was set at 0.05 for all models.[29] We corrected for 19 (SNPs) x 15 (biomarkers) = 285 tests, and our BH critical value was 0.001. For SNP genotypes that showed an association with "combined" protein concentrations in the initial analysis, the association between SNP genotypes and individual peptide fragments was also examined using linear regressions adjusting for age at sample collection, sex, and plate batch.

Results

Lysosomal biomarker measurement in PD and neurologically normal individuals

CSF samples from 173 participants with PD and 100 neurologically normal control participants were assessed (Fig. 1). GBA1 and LRRK2 mutation carriers were excluded from analysis. Among PD individuals, 99 (65.6%) were cognitively normal (PD-CN), 42 (27.8%) had mild cognitive impairment (PD-MCI), and 10 (6.6%) had dementia (PDD) at the time of CSF sampling. Others with unknown clinical cognitive evaluation are indicated by PD-NA. Demographics for all participants are shown in Table 2.

We measured 15 proteins by parallel reaction monitoring mass spectrometry (MS) in all CSF samples. These included nine lysosomal proteins as well as six additional proteins otherwise linked to neurodegenerative disease processes (Table S2). Among these analytes, adaptor-related protein complex 2 subunit beta 1 (AP2B1), lysozyme (LYZ), and ubiquitin differed significantly comparing PD and neurologically normal controls (Fig. S1). Within PD, amyloid precursor protein (APP), lysozyme (LYZ), and cathepsin D (CTSD) differed significantly comparing PD-CN, PD-MCI, and PDD (Fig. S2).

Table 2. Summary of participants.

	NC (*n* = 100)	PD (*n* = 173)	Overall (*n* = 273)
Age			
Mean (SD)	69.4 (9.25)	67.5 (8.85)	68.2 (9.03)
Median [min, max]	70.0 [40.0, 89.0]	68.0 [29.0, 89.0]	68.0 [29.0, 89.0]
Sex			
Female	58 (58.0%)	46 (26.6%)	104 (38.1%)
Male	42 (42.0%)	127 (73.4%)	169 (61.9%)

	PD-CN (*n* = 99)	PD-MCI (*n* = 42)	PDD (*n* = 10)
Age			
Mean (SD)	65.8 (7.29)	67.7 (7.40)	75.8 (6.66)
Median [min, max]	66.0 [49.0, 89.0]	67.0 [54.0, 85.0]	77.5 [65.0, 86.0]
Sex			
Female	34 (34.3%)	10 (23.8%)	0 (0%)
Male	65 (65.7%)	32 (76.2%)	10 (100%)
DRS total score			
Mean (SD)	140 (2.75)	134 (5.49)	112 (18.3)
Median [min, max]	141 [132, 144]	136 [119, 143]	120 [69.0, 126]
Age at onset			
Mean (SD)	58.3 (7.99)	58.9 (8.72)	66.2 (10.3)
Median [MIN, MAX]	59.0 [38.0, 73.0]	59.0 [36.0, 75.0]	68.5 [50.0, 82.0]

All Parkinson's disease participants (PD) and neurologically normal controls (NC) are shown in the top panel, and the 151 PD individuals with cognitive characterization by consensus diagnosis are shown grouped by cognition in the bottom panel, where PD-CN refers to cognitively normal PD, PD-MCI refers to PD with mild cognitive impairment, and PDD refers to PD with dementia. DRS Total Score refers to score on the Mattis Dementia Rating Scale-2 within 6 months of CSF sample collection. NC and PD do not differ by age or sex (Chisq = 2.560, *P* = 1.096). Age differs between PD-CN, PD-MCI, and PDD (one-way ANOVA *P* = 0.0009), with PDD individuals being older (*P* = 0.036 compared with PD-CN, and *P* = 0.0081 compared with PD-MCI in posthoc Tukey's test). DRS Total scores also differ comparing PD-CN, PD-MCI, and PDD (one-way ANOVA *P* = 0.0007), with significant differences between all pairs (*P* < 0.0001 in post hoc Tukey's test).

Systematic nomination of candidate PD risk loci

Large-scale PD GWAS have been performed previously, nominating ~90 genetic loci that associate with risk of developing PD, but functional investigations of most of these loci are lacking.[22,30] In order to systematically identify loci to test for impact on lysosomes, we began with 525 target gene-SNP pairs nominated by the MR analysis of Nalls et al. in 2019.[22] We filtered these pairs for strength of association with PD, retaining those with

P < 1e-08 (genome-wide corrected *P*-value < 0.05). We further filtered these pairs for impact on their target genes, retaining those that introduced missense mutations or impacted expression of their target genes substantially (β > 30% difference between alleles in publicly available brain data).[31] Thus, our systematic filtering approach yielded 16 PD risk SNPs for downstream evaluation.

In addition to these 16 SNPs, we tested SNPs in the *TMEM106B*, *CTSB*, and *RAB29* loci, based on previous work linking these genetic targets to lysosomal function.[12,13,16,18,20] Chosen SNPs in the *TMEM106B*, *CTSB*, and *RAB29* loci acted as expression quantitative trait loci (eQTLs) for their respective genes in multiple brain regions.[31,32] Increased expression of *TMEM106B* results in altered lysosomal acidification and degradation.[13,15] *CTSB* encodes cathepsin B, an important lysosomal protease with potential for therapeutic targeting in neurodegeneration.[33] *RAB29* encodes a Rab GTPase implicated in lysosomal stress response and possible regulation of *LRRK2*.[34,35] These three loci, while not chosen systematically, are relatively well-characterized and allowed us to investigate whether potential genetic effects on lysosomal acidification, expression of protease, or stress response might affect our lysosomal readouts. In total, our approach yielded 19 PD risk SNPs for downstream evaluation (Fig. 1).

Variants at CAMLG and ITGA8 loci associate with lysosomal biomarker levels

In a multivariable linear regression model adjusted for age, sex, and experimental batch, we assessed whether these 19 PD risk SNPs associated with CSF lysosomal biomarker levels (Table S3). We found that rs12657663 (*CAMLG*) and rs7910668 (*ITGA8*) genotypes associated with levels of at least one CSF biomarker after correction for multiple hypothesis testing (Fig. 2). Specifically, genotypes at rs7910668, a noncoding SNP with brain expression quantitative trait locus (eQTL) effects on *ITGA8*, associated with cathepsin B (CTSB) levels, while genotypes at rs12657663, a missense variant in *CAMLG*, associated with cathepsin F (CTSF), cathepsin L (CTSL), hexosaminidase B (HEXB), and tripeptidyl peptidase I (TPP1) levels in CSF (Fig. 3). Notably, although *TMEM106B*, *CTSB*, and *RAB29* are genes with known lysosomal function, SNPs at these loci did not significantly associate with any of our CSF lysosomal readouts.

Because our MS-based approach yielded measures of multiple peptides for each protein, we examined rs7910668 and rs12657663 genotypes for association with levels of each individual peptide. All three peptide fragments of CTSB showed significant associations with rs7910668 (*ITGA8*) genotypes in concordant directions,

SNP's association with CSF lysosomal biomarkers
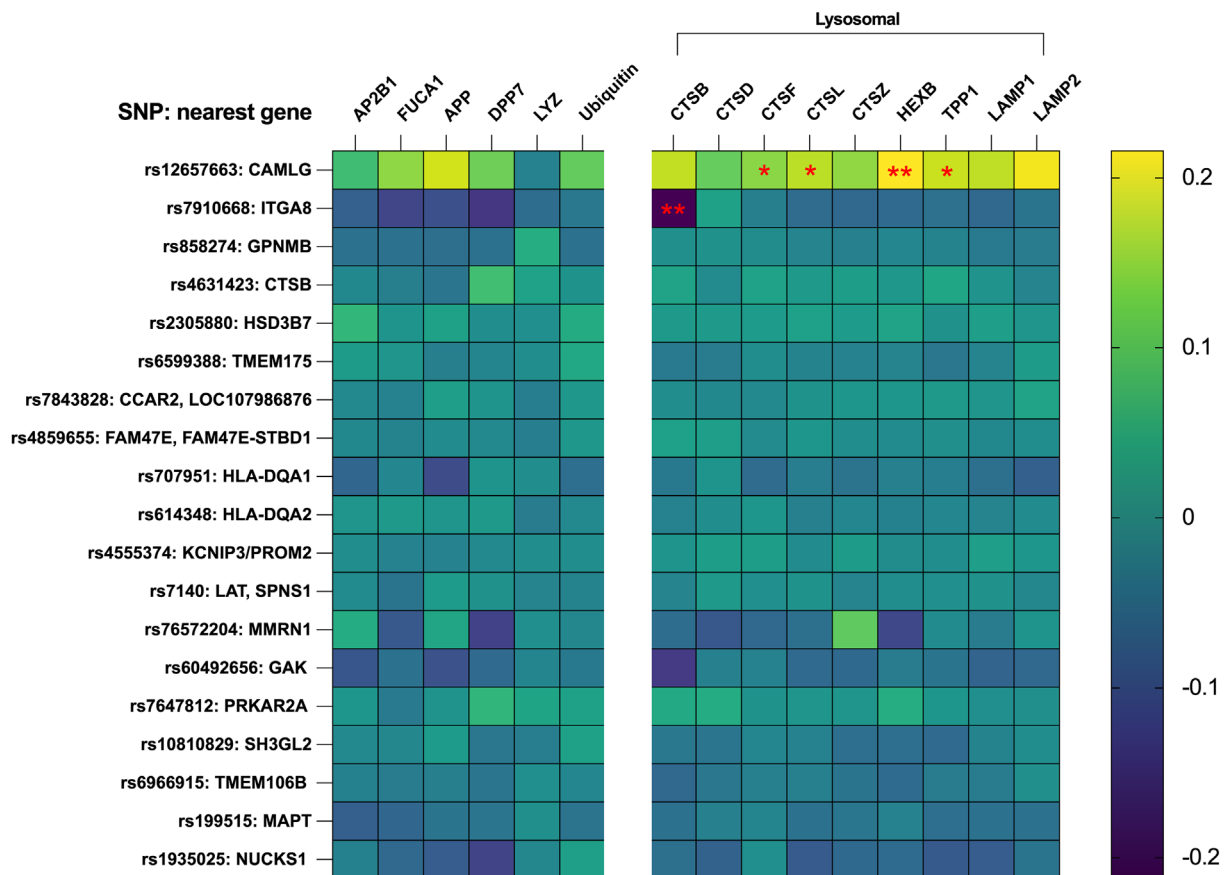
Figure 2. Associations between 19 SNPs and 15 CSF proteins. On the heatmap, each row represents one SNP, and each column represents one protein measured by MS. Colors represent the standardized beta coefficient for SNP effects on protein expression, with positive values (yellow) indicating that the PD risk allele correlates with higher expression and negative values (blue) indicating that the PD risk allele correlates with lower expression. Significant associations are indicated by asterisks *: with rs12657663 (*CAMLG*), CTSF (*P* = 0.0165), CTSL (*P* = 0.0201), HEXB (*P* = 0.0036) and TPP1 (*P* = 0.0320); with rs7910668 (*ITGA8*), CTSB (*P* = 0.0074). Benjamini–Hochberg (BH) adjusted *P*-values are indicated by asterisks (**P* < 0.05, ***P* < 0.01) and were calculated through linear regression adjusted for age, sex, and plate batch.

with the PD risk allele at this SNP (A) correlating with lower CSF CTSB (Fig. S3A). For rs12657663 (*CAMLG*), all individual peptide fragments for CTSF, CTSL, HEXB, and TPP1 showed significant associations with genotype: the major allele (G) for rs12657663, associated with PD risk, also associated with higher levels of these lysosomal readouts (Fig. S3B).

## CAML protein characterization in human brain

rs12657663 is a missense variant in *CAMLG* that changes amino acid 78 from valine to isoleucine (V78I, Fig. S4) in the encoded protein Calcium modulating ligand (CAML). While the role of CAML in insertion of tail-anchored proteins has been extensively characterized at the cell

biological level,[36,37] CAML function and expression in the human brain is largely unknown.

We characterized CAML expression in six brain regions (midbrain, thalamus, hippocampus, frontal cortex, temporal cortex, and cerebellum), finding that CAML is predominantly expressed in neurons throughout all regions evaluated (Fig. 4). Because the Human Protein Atlas shows low to no expression of Integrin subunit alpha 8 (the protein product of *ITGA8*) in the brain, we did not perform similar brain studies for this protein.

CAML protein expression across neurodegenerative diseases in hippocampus

Differences noted incidentally in our regional survey of 12 PD and 12 NC cases led us to hypothesize that CAML
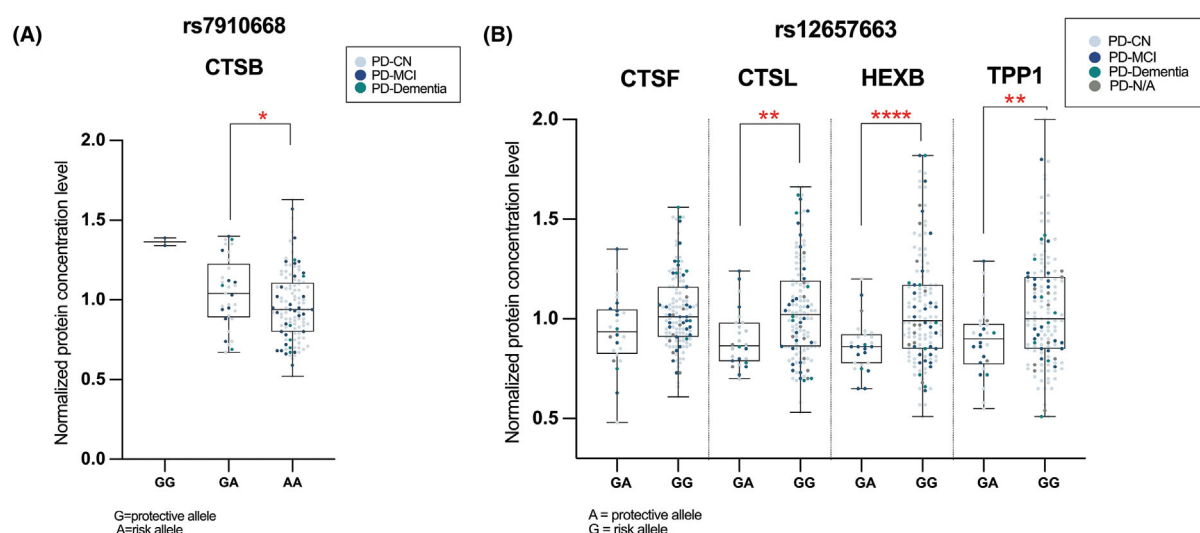
Figure 3. Associations of rs7910668 and rs12657663 genotypes with CSF lysosomal proteins. (A) Genotypes at rs7910668 (*ITGA8*) are shown on the x-axis, and CSF CTSB levels on the y-axis for $n = 154$ individuals. (B) Genotypes at rs12657663 (*CAMLG*) are shown on the x-axis, and CSF levels for cathepsin F (CTSF), cathepsin L (CTSL), hexosaminidase B (HEXB), and tripeptidyl peptidase I (TTP1) on the y-axis for $n = 156$ individuals. Some proteins did not have reliable measures for all individuals, accounting for slight differences in numbers. Benjamin–Hochberg (BH) adjusted P-values comparing protein levels across different genotypes (without adjustment for covariants) are indicated by asterisks (*$P < 0.05$, **$P < 0.01$, **** $P < 0.001$). CSF lysosomal protein concentrations were normalized to an average of 1.

could be more highly expressed in PD cases. To better understand whether disease status impacts CAML expression in the brain, we assessed CAML expression in the hippocampus of 56 neuropathological cases (9 NC, 17 PD without concomitant Alzheimer's disease (PD+/AD−), 11 PD with concomitant AD (PD+/AD+), 7 AD without concomitant Lewy body disease (AD+/LBD−), 6 AD with concomitant LBD (AD+/LBD+), and 6 amyotrophic lateral sclerosis (ALS) cases). As shown in Fig. 5, cases with Lewy body pathology (PD+/AD−, PD+/AD+, AD+LBD+) demonstrated visibly increased expression of CAML, compared with cases without Lewy body pathology. Quantification of CAML staining across all cases corroborated our qualitative assessment ($n = 34$ for Lewy pathology cases versus $n = 22$ for cases without Lewy body pathology, $P < 0.0001$ through pairwise comparison). Moreover, one-way ANOVA indicated significant differences in CAML staining across groups ($P = 0.0018$), with pairwise differences for NC versus PD+/AD− ($P = 0.0141$), PD+/AD− versus AD+/LBD− ($P = 0.0447$), and AD+/LBD− versus AD+/LBD+ ($P = 0.0002$).

Discussion

In the present work, we associate GWAS-derived PD risk SNPs with CSF lysosomal protein levels to ask whether PD risk variants implicate lysosomal function generally, and to find specific variants and target genes that impact lysosomes. To our surprise, relatively few PD risk SNPs—

and none in loci selected for previously reported lysosomal functions—associated with the lysosomal proteins measured in our study. However, two loci—rs7910668 near *ITGA8* and rs12657663 in *CAMLG*—demonstrated protein quantitative trait locus (pQTL) effects on one or more CSF lysosomal proteins. In addition, CAML is expressed in neurons in multiple brain areas, with more robust staining in postmortem brain samples with Lewy body pathology. Taken together, our work newly implicates two PD risk loci in lysosomal function and suggests that genetic variants in *CAMLG* act in neuronal lysosomes.

rs7910668 is a noncoding SNP on chromosome 10 near *ITGA8*, for which it acts as an eQTL. PD risk variants at rs7910668 associate with lower *ITGA8* expression, and with lower CTSB levels in CSF.[22,32] Variants in the *CTSB* locus that decrease *CTSB* expression have previously been reported to increase risk for PD, including among *GBA1* variant carriers.[38–40] Moreover, induced pluripotent stem cell-derived neurons (iPSC-N) from *GBA1* N370S mutation carriers with PD have decreased CTSB protein expression compared to iPSC-N from PD individuals without *GBA1* mutations.[38,41] *CTSB* and *ITGA8* are on different chromosomes, and rs7910668 is in linkage equilibrium with PD-associated SNPs (*e.g.*, rs1293298) at the *CTSB* locus. Considering these data in light of our present findings, we hypothesize that variants in rs7910668 increase risk for PD by lowering CTSB expression. Whether this effect occurs mechanistically through
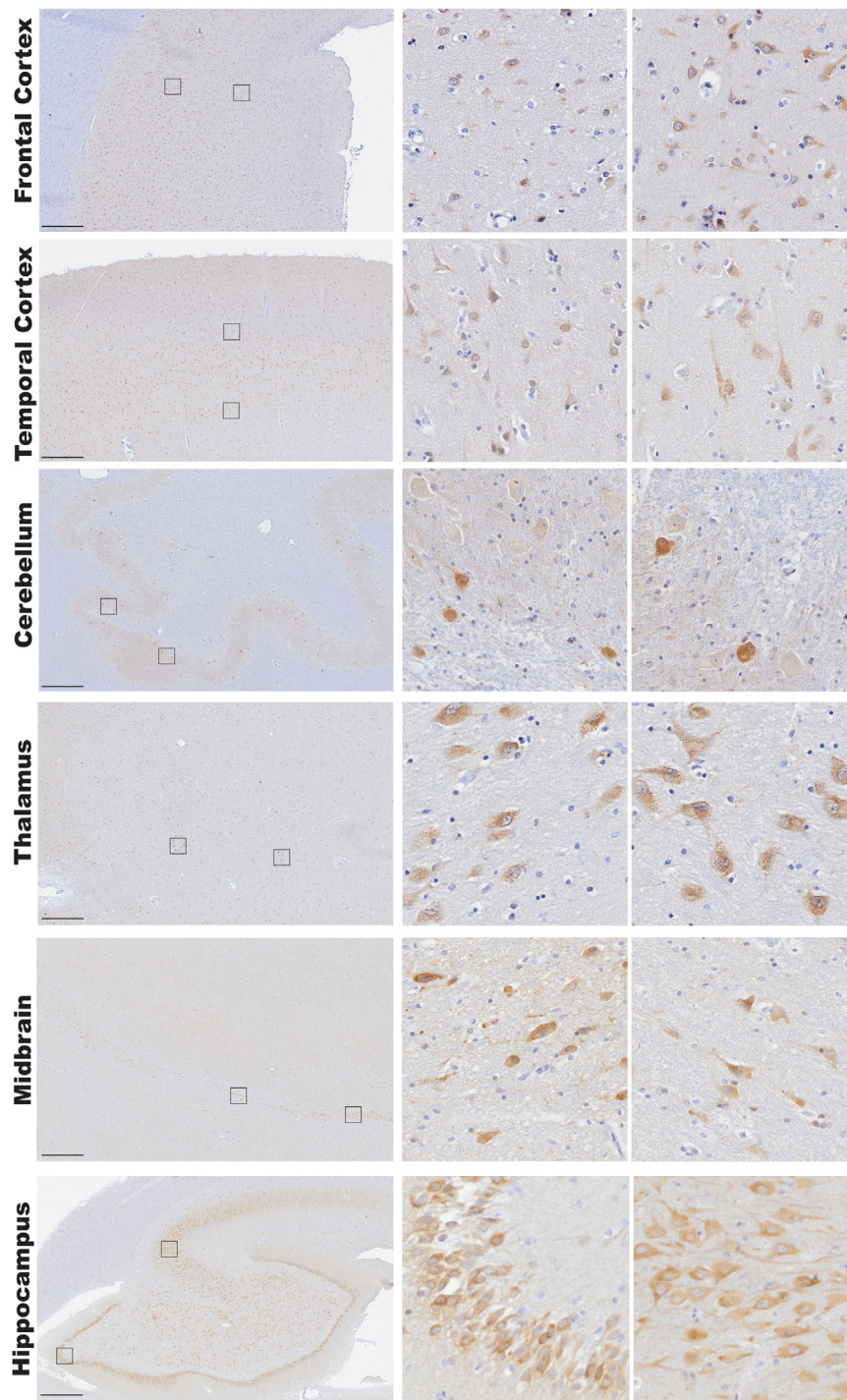
Figure 4. Representative immunohistochemical staining of CAML expression in multiple brain regions. CAML is expressed in neurons across all regions assessed, including frontal cortex, temporal cortex, cerebellum, thalamus, midbrain, and hippocampus. Sections were stained with anti-CAML antibody (NovusBio, 1:750), and images captured at 20X with an Olympus VS200 Slide Scanner. Twelve PD and 12 NC cases were evaluated: case information is available in Table S1. Insets represent 20X zoom. Scale bar = 500 μm.
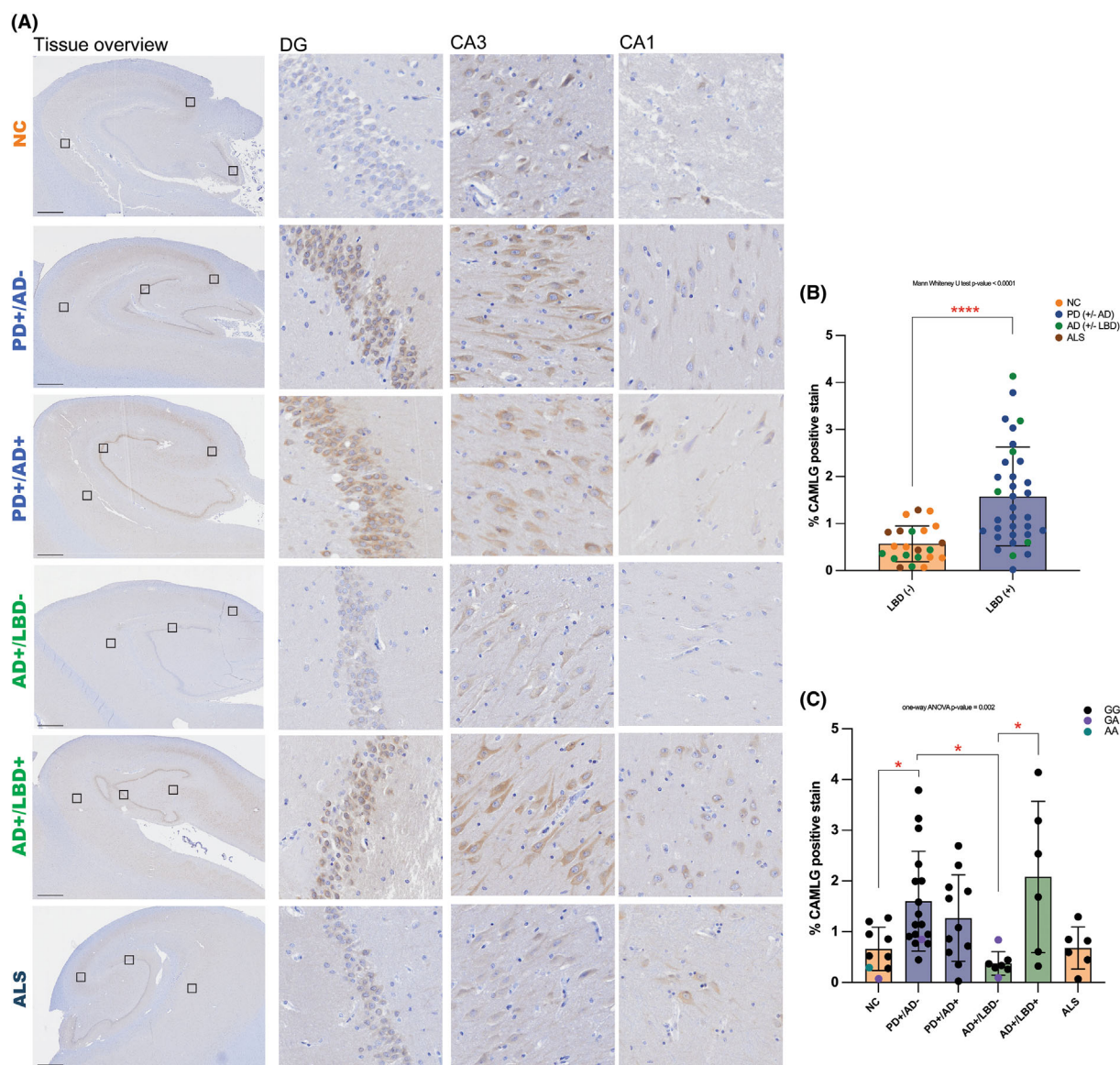
Figure 5. Hippocampal CAML expression across a spectrum of neurodegenerative diseases. (A) Representative immunohistochemical staining of hippocampal sections from NC ($n = 9$ assessed), PD+/AD− ($n = 17$ assessed), PD+/AD+ ($n = 11$ assessed), AD+/LBD− ($n = 7$ assessed), AD+/LBD+ ($n = 6$ assessed), and ALS ($n = 6$ assessed) cases. Images were acquired at 20X, and insets of dentate gyrus (DG), CA3, and CA1 subregions are magnified 20X. Scale bar =500um. (B) CAML expression was compared in cases without Lewy body pathology (shown as "LBD (−)," $n = 22$) versus cases with LBD pathology (shown as "LBD (+)," $n = 34$) pathology cases. Mann–Whitney test was performed to compare the two groups and showed significant differences in "LBD (−)" vs. "LBD (+)" cases. (****, $P < 0.0001$). (C) CAML expression was quantified across all 56 cases. One-way ANOVA demonstrated significant differences across all groups ($P = 0.0018$), with significant differences between pairs indicated by asterisks (* Tukey's HSD method $P < 0.05$).

activities of integrin subunit alpha 8 (encoded by *ITGA8* and known to regulate neurite outgrowth[42]) remains unclear.

The missense variant in *CAMLG* rs12657663 impacts CSF measures of multiple lysosomal enzymes, including cathepsins F and L, the peptidase TTP1, and HEXB (which catalyzes the degradation of the ganglioside GM2)

[46–48].[43,44] PD risk variants in *CAMLG* associate with higher levels of all four lysosomal enzymes, suggesting that CAML may act in PD pathogenesis by compromising lysosomal function, leading to "overflow" of lysosomal protein signal into the CSF.[26,32] Whether *CAMLG* variants act through amino acid change (since rs12657663 results in V78I change) or by modulating CAML expression

(since rs12657663 also associates with *CAMLG* RNA expression) remains to be seen. Of note, the PD risk variant at rs12657663 associates with lower *CAMLG* expression, while LBD cases in our neuropathological analysis showed more robust CAML staining.

The major strengths of our study lie in its extensive use of large, orthogonal human sample sets and its unbiased approach to understanding PD risk variants. Specifically, our CSF biomarker characterization included 273 individuals with and without PD, and our neuropathological characterization of CAML incorporated 56 postmortem cases. Moreover, PD genetic studies were performed (by other investigators) in cohorts that were distinct from those used in the biofluid biomarker study, which also did not overlap with our neuropathological cohort. Few functional genomic studies have incorporated multiple different cohorts of this size, but the size of our cohorts increases statistical power, and the extensive use of orthogonal human data increases confidence in our findings. At the same time, few biomarker studies have incorporated an unbiased approach to investigating GWAS-derived genetic risk variants—most focus on the impact of a few candidate genetic mutations or variants on a biomarker of interest.[45–47] In contrast, we began with all known PD GWAS-derived variants, filtered them systematically through a combination of causal inference analyses and prioritization by effect size and strength of statistical evidence, in order to develop a global view of the impact of PD genetic risk on lysosomal biomarker signatures.

Limitations of our study should be considered along with its strengths. For example, the tested genetic variants likely only represent a subset of variants implicated in lysosomal function and in PD risk, because studies uncovering PD genetic risk variants, as well as methods for linking them to potential target genes, are perpetually improved and updated. In this vein, we note that, in contrast to our nomination of the other 16 PD risk variants for study using unbiased rules, our decision to test variants in *TMEM106B*, *CTSB*, and *RAB29* was subjective. We chose these genetic loci for evaluation based on their extensive functional characterization, but certainly other loci could have been chosen with reasonable justification. Second, the tested lysosomal proteins are not the only possible readouts for lysosomal pathways. Rather, they represent a suite of MS-based assays that have been previously well-validated and cover a range of possible lysosomal functions. Third, the neuropathological characterization presented here is based on immunohistochemical staining, which is a semi-quantitative method, and the 56-case analysis focused on the hippocampus. That said, the stark increases identified by an automated image-analysis pipeline in CAML staining of cases with

vs, without LBD argues for true disease-associated differences. Fourth, while our study measured multiple lysosomal proteins in CSF from a large number of PD individuals, we are still underpowered to detect smaller effects. Indeed, for 80% power to detect an effect size of 0.2 by linear regression at an unadjusted alpha of 0.05, a sample size of 199 (greater than the current study size of 173) is needed. For 80% power to detect an effect size of 0.2 with the Benjamini–Hochberg corrected *P*-values used in our study, a sample size of 433 is required. Thus, our results represent the most robust, rather than the only, relationships between PD genetic risk factors and lysosomal phenotypes. Fifth, it is presently uncertain to what extent CSF concentrations of the measured lysosomal proteins reflect neuronal lysosomal function.[48] Additionally, the data presented are correlational, so it is difficult to determine the directions of causality. For example, does CAML alter lysosomal function that then impacts levels of CSF lysosomal enzymes, as we have hypothesized? Or does CAML directly impact levels of the cathepsins, TTP1, and HEXB through its cell biological function? In this context, it is worth noting that while effects of genetic variation at rs12657663 on CAML have not been studied, this protein is relatively well characterized at the cell biological level. Specifically, CAML is the mammalian homolog of Get2, known to mediate post-translational insertion of tail-anchored proteins into the endoplasmic reticulum as part of the Guided Entry of Tail-anchored protein (GET) system.[36,37,49–52] Downstream repercussions of changes in tail-anchored protein insertion in neurons are likely to be significant: mutations in the GET protein WRB, for example, result in neuronal phenotypes.[53,54] Thus, future validation studies investigating the consequences of changes in *CAMLG* expression, and of V78I, on neuronal lysosomal function and handling of "proteinopathies" such as misfolded aSyn in model systems such as neuronal culture or transgenic animals, are easy to envision and would be a valuable addition to the work presented here.

These limitations notwithstanding, our study adopted an unbiased approach to PD genetic risk loci, allowing the data to tell us which ones might function in lysosomal pathways. In this way, we pinpointed two loci—within *CAMLG* and near *ITGA8*—with strong evidence for impact on lysosomal function in human PD. We note that our study differs slightly from the conventional functional genomics analysis in that we probe a specific function—lysosomal function—based on extensive evidence for its involvement in PD pathogenesis. We argue that such a targeted approach may be particularly fruitful if the goal is to develop testable hypotheses for downstream mechanistic investigation. Detailed mechanistic investigation, in turn, is needed if we are to develop viable

therapeutic targets from the wealth of genomic data in neurodegenerative disease.

Author Contributions

Acknowledgements

Funding Information

Conflicts of Interests

The following authors declare that they have no competing interests: ACP, Hanwen Z, TT, JR, JS, EB, Ann B, SS, Adama B, MCC, ES, VVD, KB, DW, and EL.

Henrik Z. has served on scientific advisory boards and/or as a consultant for multiple commercial entities, but these entities were not involved in the study.

Data Availability Statement

R scripts for analysis are included in the supplement and also shared via the Chen-Plotkin lab GitHub (https://github.com/ChenPlotkinLab/CAMLG). All relevant data are included in the manuscript or in supplementary files, but raw data can be provided upon request. Samples collected are available for request through a biospecimen request access committee; inquiries should be addressed to the corresponding author.

References

1. Werner Poewe, Klaus Seppi, Tanner CM, Halliday GM, et al. Parkinson disease. Nat Rev Dis Primers. 2017;3 (17013).
2. Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. J Neural Transm (Vienna). 2017;124(8):901-905.
3. Kowal SL, Dall TM, Chakrabarti R, Storm MV, Jain A. The current and projected economic burden of Parkinson's disease in the United States. Mov Disord. 2013;28(3):311-318.
4. Alba Navarro-Romero MM, Martinez-Vicente M. The emerging role of the lysosome in Parkinson's disease. Cells. 2020;9(11)
5. Carceles-Cordon M, Weintraub D, Chen-Plotkin AS. Cognitive heterogeneity in Parkinson's disease: a mechanistic view. Neuron. 2023;111:1531-1546.
6. Migdalska-Richards A, Schapira AH. The relationship between glucocerebrosidase mutations and Parkinson disease. J Neurochem. 2016;139:77-90.
7. Do J, McKinney C, Sharma P, Sidransky E. Glucocerebrosidase and its relevance to Parkinson disease. Mol Neurodegener. 2019;14:36.
8. Gabbert C, Schaake S, Lüth T, et al. GBA1 in Parkinson's disease: variant detection and pathogenicity scoring matters. BMC Genomics. 2023;24:322.
9. Mazzulli Joseph R, Xu Y-H, Sun Y, et al. Gaucher disease Glucocerebrosidase and α-Synuclein form a bidirectional pathogenic loop in Synucleinopathies. Cell. 2011;146 (1):37-52.
10. Gegg ME, Burke D, Heales SJ, et al. Glucocerebrosidase deficiency in substantia nigra of parkinson disease brains. Ann Neurol. 2012;72(3):455-463.
11. Tropea TF, Amari N, Han N, et al. Whole clinic research enrollment in Parkinson's disease: the molecular integration in neurological diagnosis (MIND) study. J Parkinsons Dis. 2021;11(2):757-765.
12. Tropea TF, Mak J, Guo MH, et al. TMEM106B effect on cognition in Parkinson disease and frontotemporal dementia. Ann Neurol. 2019;85(6):801-811.
13. Alice S, Chen-Plotkin TLU, Gallagher MD, et al. TMEM106B, the risk gene for frontotemporal dementia, is regulated by the microRNA-132/212 cluster and affects Progranulin pathways. J Neurosci. 2012;32 (33):11213-11227.
14. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. Am J Hum Genet. 2018;102 (5):717-730.
15. Busch JI, Unger TL, Jain N, Tyler Skrinak R, Charan RA, Chen-Plotkin AS. Increased expression of the frontotemporal dementia risk factor TMEM106B causes C9orf72-dependent alterations in lysosomes. Hum Mol Genet. 2016;25(13):2681-2697.
16. Chang A, Xiang X, Wang J, et al. Homotypic fibrillization of TMEM106B across diverse neurodegenerative diseases. Cell. 2022;185(8):1346-1355.e15.
17. Perneel J, Neumann M, Heeman B, et al. Accumulation of TMEM106B C-terminal fragments in neurodegenerative disease and aging. Acta Neuropathol. 2023;145(3):285-302.
18. Kim MJ, Jeong H, Krainc D. Lysosomal ceramides regulate cathepsin B-mediated processing of saposin C and glucocerebrosidase activity. Hum Mol Genet. 2022;31 (14):2424-2437.
19. Yang SY, Gegg M, Chau D, Schapira A. Glucocerebrosidase activity, cathepsin D and monomeric

α-synuclein interactions in a stem cell derived neuronal model of a PD associated GBA1 mutation. Neurobiol Dis. 2020;134:104620.

20. Cunningham LA, Moore DJ. Endosomal sorting pathways in the pathogenesis of Parkinson's disease. Prog Brain Res. 2020;252:271-306.

21. Purlyte E, Dhekne HS, Sarhan AR, et al. Rab29 activation of the Parkinson's disease-associated LRRK2 kinase. EMBO J. 2018;37(1):1-18.

22. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019;18 (12):1091-1102.

23. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry. 1992;55(3):181-184.

24. Pigott K, Rick J, Xie SX, et al. Longitudinal study of normal cognition in Parkinson disease. Neurology. 2015;85 (15):1276-1282.

25. Toledo JB, Van Deerlin VM, Lee EB, et al. A platform for discovery: the University of Pennsylvania Integrated Neurodegenerative Disease Biobank. Alzheimers Dement. 2014;10(4):477-484.e1.

26. Sjödin S, Brinkmalm G, Öhrfelt A, et al. Endo-lysosomal proteins and ubiquitin CSF concentrations in Alzheimer's and Parkinson's disease. Alzheimers Res Ther. 2019;11:82.

27. Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9(6):477-485.

28. Chahine LM, Qiang J, Ashbridge E, et al. Clinical and biochemical differences in patients having Parkinson disease with vs without GBA mutations. JAMA Neurol. 2013;70(7):852-858.

29. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Methodol. 1995;57(1):289-300.

30. Chang D, Nalls MA, Hallgrímsdóttir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat Genet. 2017;49 (10):1511-1516.

31. Aguet F, Brown AA, Castel SE, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550:204-213.

32. Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45:580-585.

33. Sharma A, Swetha R, Bajad NG, et al. Cathepsin B-A neuronal death mediator in Alzheimer's disease leading to neurodegeneration. Mini Rev Med Chem. 2022;22 (15):2012-2023.

34. Komori T, Kuwahara T, Fujimoto T, et al. Phosphorylation of Rab29 at Ser185 regulates its

localization and role in the lysosomal stress response in concert with LRRK2. J Cell Sci. 2023;136(14):jcs261003

35. Zhu H, Tonelli F, Turk M, Prescott A, Alessi DR, Sun J. Rab29-dependent asymmetrical activation of leucine-rich repeat kinase 2. Science. 2023;382(6677):1404-1411.

36. Yamamoto Y, Sakisaka T. Molecular machinery for insertion of tail-anchored membrane proteins into the endoplasmic reticulum membrane in mammalian cells. Mol Cell. 2012;48(3):387-397.

37. Vilardi F, Stephan M, Clancy A, Janshoff A, Schwappach B. WRB and CAML are necessary and sufficient to mediate tail-anchored protein targeting to the ER membrane. PLoS One. 2014;9(1):e85033.

38. Milanowski LM, Hou X, Bredenberg JM, et al. Cathepsin B p.Gly284Val variant in Parkinson's disease pathogenesis. Int J Mol Sci. 2022;23(13):7086.

39. Drobny A, Prieto Huarcaya S, Dobert J, et al. The role of lysosomal cathepsins in neurodegeneration: mechanistic insights, diagnostic potential and therapeutic approaches. Biochim Biophys Acta Mol Cell Res. 2022;1869:119243.

40. Tsujimura A, Taguchi K, Watanabe Y, et al. Lysosomal enzyme cathepsin B enhances the aggregate forming activity of exogenous α-synuclein fibrils. Neurobiol Dis. 2015;73:244-253.

41. Blauwendraat C, Reed X, Krohn L, et al. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. Brain. 2020;143(1):234-248.

42. Varnum-Finney B, Venstrom K, Muller U, et al. The integrin receptor alpha 8 beta 1 mediates interactions of embryonic chick motor and sensory neurons with tenascin-C. Neuron. 1995;14(6):1213-1222.

43. Cordeiro P, Hechtman P, Kaplan F. The GM2 gangliosidoses datanases: allelic variation at the HEXA, HEXB, and GMWA gene loci. Genet Med. 2000;2:319-327.

44. Kuil LE, López Martí A, Carreras Mascaro A, et al. Hexb enzyme deficiency leads to lysosomal abnormalities in radial glia and microglia in zebrafish brain development. Glia. 2019;67(9):1705-1718.

45. Cruchaga C, Graff C, Chiang HH, et al. Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels. Arch Neurol. 2011;68(5):581-586.

46. Saddiki H, Fayosse A, Cognat E, et al. Age and the association between apolipoprotein E genotype and Alzheimer disease: a cerebrospinal fluid biomarker-based case-control study. PLoS Med. 2020;17(8):e1003289.

47. Piccio L, Deming Y, Del-Águila JL, et al. Cerebrospinal fluid soluble TREM2 is higher in Alzheimer disease and associated with mutation status. Acta Neuropathol. 2016;131(6):925-933.

48. Gobom J, Brinkmalm A, Brinkmalm G, Blennow K, Zetterberg H. Alzheimer's disease biomarker analysis using targeted mass spectrometry. Mol Cell Proteomics. 2024;23(2):100721.

49. Schuldiner M, Metz J, Schmid V, et al. The GET complex mediates insertion of tail-anchored proteins into the ER membrane. Cell. 2008;134(4):634-645.

50. Asseck LY, Mehlhorn DG, Monroy JR, et al. Endoplasmic reticulum membrane receptors of the GET pathway are conserved throughout eukaryotes. Proc Natl Acad Sci USA. 2021;118(1):e2017636118.

51. Mariappan M, Mateja A, Dobosz M, Bove E, Hegde RS, Keenan RJ. The mechanism of membrane-associated steps in tail-anchored protein insertion. Nature. 2011;477 (7362):61-66.

52. Yamamoto Y, Sakisaka T. The emerging role of calcium-modulating cyclophilin ligand in posttranslational insertion of tail-anchored proteins into the endoplasmic reticulum membrane. J Biochem. 2015;157(6):419-429.

53. Vogl C, Panou I, Yamanbaeva G, et al. Tryptophan-rich basic protein (WRB) mediates insertion of the tail-anchored protein otoferlin and is required for hair cell exocytosis and hearing. EMBO J. 2016;35(23):2536-2552.

54. Daniele LL, Emran F, Lobo GP, Gaivin RJ, Perkins BD. Mutation of wrb, a component of the guided entry of tail-anchored protein pathway, disrupts photoreceptor synapse structure and function. Invest Ophthalmol Vis Sci. 2016;57(7):2942-2954.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1:
Table S2:
Table S3: