omicSynth: an Open Multi-omic Community Resource for Identifying Druggable Targets across Neurodegenerative Diseases

Chelsea X. Alvarado[1,2], Mary B. Makarious[3,4,5], Dan Vitale[2], Mathew J. Koretsky[1], Sara Bandres-Ciga[1], Hirotaka Iwaki[1,2,3], Kristin Levine[1,2], Andrew Singleton[1,3], Faraz Faghri[1,2,3], Mike A. Nalls[1,2,3], Hampton L. Leonard[1,2,3,6]*

Correspondence

Hampton Leonard

Center for Alzheimer's and Related Dementias

National Institute on Aging, Intramural Research Program at the National Institutes of Health

Building T44

Bethesda, MD 20892, USA

E-Mail: hampton@datatecnica.com

Affiliations

1. Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA, 20814
2. Data Tecnica International, Washington, DC, USA, 20037
3. Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA, 20814
4. Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK, WC1N 3BG
5. UCL Movement Disorders Centre, University College London, London, UK, WC1N 3BG
6. German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany

Keywords

Neurodegeneration, druggability, Summary-data-based Mendelian Randomization, SMR

Abstract

Treatments for neurodegenerative disorders remain rare, although recent FDA approvals, such as Lecanemab and Aducanumab for Alzheimer's Disease, highlight the importance of a mechanistic approach in creating disease modifying therapies. As a large portion of the global population is aging, there is an urgent need for therapeutics that

can stop disease progression and eliminate symptoms. In this study, we create an open framework and resource for evidence based identification of therapeutic targets for neurodegenerative disease. We use Summary-data-based Mendelian Randomization to identify genetic targets for drug discovery and repurposing. In parallel, we provide mechanistic insights into disease processes and potential network-level consequences of gene-based therapeutics. We identified 116 Alzheimer's disease, 3 amyotrophic lateral sclerosis, 5 Lewy body dementia, 46 Parkinson's disease, and 9 Progressive supranuclear palsy target genes passing multiple test corrections ($p_{SMR_multi} < 2.95E-06$ and $p_{HEIDI} > 0.01$). We created a therapeutic scheme to classify our identified target genes into strata based on druggability and approved therapeutics - classifying 41 *novel* targets, 3 *known* targets, and 115 *difficult* targets. Our *novel* class of genes provides a springboard for new opportunities in drug discovery, development and repurposing in the pre-competitive space. We also provide a user-friendly web platform to help users explore potential therapeutic targets for neurodegenerative diseases, decreasing activation energy for the community [https://nih-card-ndd-smr-home-syboky.streamlit.app/].

Introduction

Currently, there are few approved disease-modifying therapeutics available to those with neurodegenerative diseases (NDD), the most recent being Lecanamab for the treatment of Alzheimer's disease [1]. NDD such as Alzheimer's disease (AD), Parkinson's disease (PD), Amyotrophic lateral sclerosis (ALS), Lewy body dementia (LBD), Frontotemporal lobar degeneration (FTLD), and progressive supranuclear palsy (PSP) are diseases caused by progressive nerve cell degeneration that result in a loss of cognition and/or motor function [2]. The World Health Organization (WHO) expects dementia diagnoses alone to reach 78 million in 2030 and 139 million in 2050. Without disease modifying therapies, the devastating health, social, and economic impacts of dementia and related NDD will be catastrophic [3]. There exists a clear need to generate new understanding of the basis of NDD and to identify rational therapeutic targets for these diseases; this need will require not only the generation of new data, but also the development and deployment of rapid, open, and transparent tools with which to analyze these data and identify targets.

Drugs that are supported by genetic or genomic data frequently outperform those without such evidence in clinical trials. Over two-thirds of the Food and Drug Administration (FDA) approved drugs in 2021 were supported by genetic or genomic evidence [4]. Therapeutics with genetically supported target mechanisms are twice as likely to pass a trial phase as those without supporting genetic data [5]. Given the importance of anchoring therapeutic targets to a disease mechanism substantiated by genetic evidence, we have set out to create a resource that would identify therapeutic targets in the neurodegenerative disease space using large scale genetic and genomic data, and to make that resource dynamic, open, and accessible.

To accomplish this, we combined genetic data, primarily in the form of genome-wide association studies (GWAS) from population-scale resources including data on millions of samples across multiple neurodegenerative diseases, and genomic data from tissue banks in the form of quantitative trait loci (QTL) studies taken from methylation, gene expression, chromatin, and proteomics assays. To integrate these sources of genetic and genomic evidence we used summary-data-based Mendelian randomization (SMR), a post-GWAS method that facilitates functional inferences relating to disease etiology, statistically connecting disease risk from GWAS to QTL data in relevant cell types as a means of identifying potential mechanisms. Additionally, in an attempt to address the lack of multi ancestry data in NDD research, we incorporated limited multiancestry eQTL data into our SMR analyses in search of significant and meaningful functional inferences about the similarities or differences between ancestral populations.

In this work, we prioritize identified genes as therapeutic targets of interest, leveraging known druggability and product market information, into three categories: "*novel*", "*known*", and "*difficult*" druggable targets (**Figure 1a-b**). We define *novel* targets as genes with significant functional inferences in relevant tissue and cell types in druggable regions of the genome that are not currently targeted by disease-specific therapeutics and should be prioritized in future repurposing studies. *Known* targets include genes within relevant tissue and cell types that have documented significant functional inferences, but are impacted by a known drug that specifically targets any NDD. Lastly, *difficult* targets are significant genes from the SMR analyses that are not in regions of the genome currently annotated as druggable. For all novel targets, we also examined up- and down-stream in the target network to identify companion genes that could also be useful as therapeutic targets for the primary gene of interest. Potential upstream and downstream effects on targeting these genes for therapeutic intervention were also provided based on network memberships and toxicity within these networks was evaluated by way of evaluating liver QTLs within the network. To make this data more accessible to the research community we have also developed a companion web-based platform to accompany this research, further decreasing activation energy for some biotech community members to explore targets [https://nih-card-ndd-smr-home-syboky.streamlit.app/].

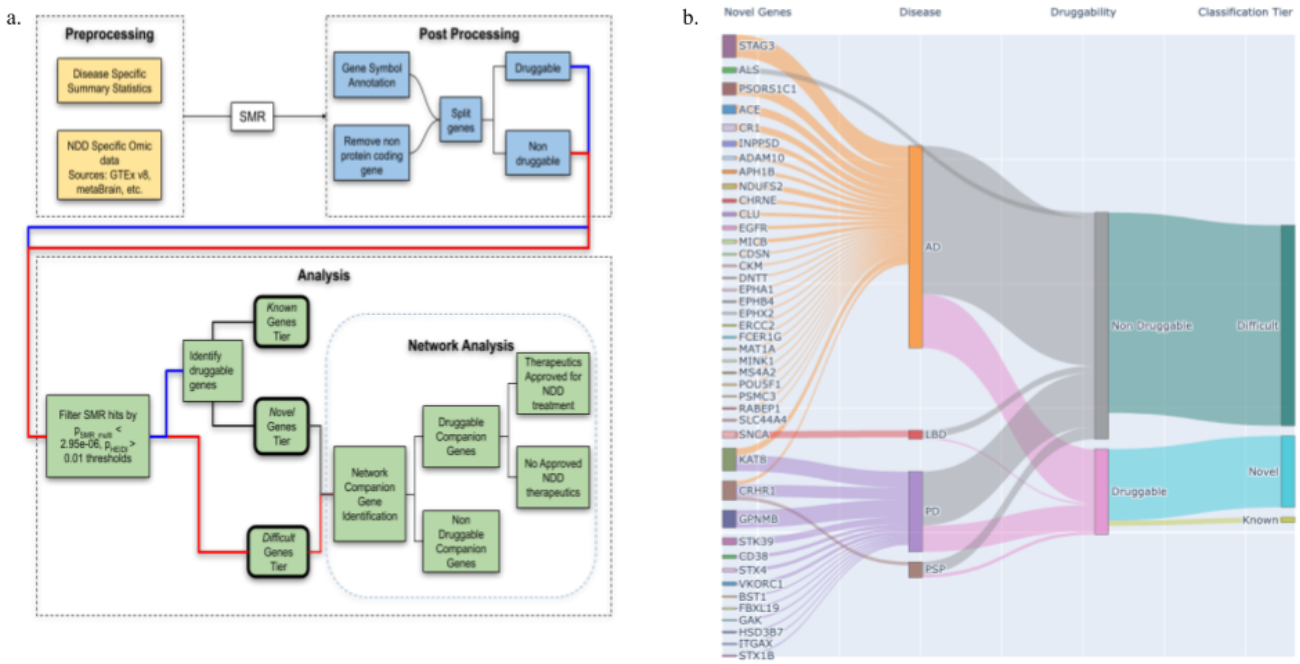

Figure 1: **a. Workflow Summary** | Graphical representation of general workflow used in conducting our analyses. NDD = Neurodegenerative Disease, SMR = Summary-data-based Mendelian Randomization **b. Graphical Summary of Results |** Sankey plot depicting the flow of candidate genes into their respective tier. On the left, we highlight novel genes but the remainder of the plot visualizes all 159 candidate genes regardless of final classification tier.

Methods

GWAS Summary Statistics

Genome-wide Association Studies (GWAS) summary statistics for each of the six NDDs highlighted in our study were used to obtain SNPs that served as instrumental variables in the Mendelian randomization pipeline. GWASs used are the latest and/or largest for each corresponding disease: Bellenguez et al., 2022 for AD (n = 788,989), Chia et al., 2021 for LBD (n = 7,372), Hoglinger et al 2011 for PSP (n = 4,361), Nalls et al., 2019 for PD (n = 1,456,306), Nicolas et al., 2018 for ALS (n = 80,610), and Pottier et al., 2019 for FTLD (n = 1,355)[6-11]. All GWAS summary statistics were lifted over, as needed, to hg19 (GRCh37) using University of California, Santa Cruz's liftOver command line tool [12].

x-QTL Summary Statistics

Expression quantitative trait loci (eQTL), protein quantitative trait loci (pQTL), chromatin quantitative trait loci (caQTL) and methylation quantitative trait loci (mQTL) were used as the outcome variables in the Mendelian randomization (MR) analyses. eQTLs are genetic loci that explain the variation in mRNA expression levels. Cis-eQTL, eQTLs that act on local genes, data makes up the majority of all x-QTL data used for our study due to the

volume of publicly available data sources. All eQTL and mQTL data obtained, except from the sources eQTLgen, metaBrain and Zeng, et al. (multi ancestry), were already in SMR format and obtained from the Yang Lab's Data Resource page [13-15]. The eQTL sources from the Yang Lab include Genotype-Tissue Expression (GTEx) project v8 release, PsychENCODE, and BrainMeta v1 (formerly brain-eMeta) [16-18]. The tissues used varied by data source but consisted of NDD-related tissues, which we have defined in this study as brain, nerve, muscle, blood, and liver tissues. Liver was included due to its role in metabolizing medications, toxicity, and potential impacts on clinical trial progress [19].

mQTLs are genetic variants that affect methylation patterns of CpG sites. mQTL data sources include Brain-mMeta and McRae et al., 2018 which are derived from blood tissues [18,20]. We additionally included caQTLs - caQTLs alter traits by modifying chromatin structure - data from Bryois et al., obtained from the Yang Lab, in our analysis but it did not return any significant associations [21]. Blood tissues have been shown to have high correlation in expression levels with brain tissues, allowing blood tissues to provide a gain of power and ease of use in biomarker studies due to the relative ease of availability of this tissue [18]. All genome positions are mapped to the human reference genome build hg19 (GRCh37).

pQTLs are genetic variants that regulate protein expression levels. Similarly to eQTLs and mQTLs, pQTLs can be used as our exposure variable. We obtained pQTL summary statistics data from Yang et al., 2021. The pQTLs are from plasma, brain and cerebrospinal fluid tissues (CSF) from participants with and without AD. Samples are on human reference genome build hg19 (GRCh37). More details on the samples and methods used can be found in the original manuscript [22]. pQTLs were nominated for inclusion for SMR analysis if they were significant ($p < 0.05$) in at least one of the three tissues per the original manuscript. We included 453 unique pQTLs across the three tissues; 223 pQTLs from CSF, 159 from plasma, and 77 from the brain into our analyses.

Gene Co-expression Networks

Data was obtained from the Open Targets and Gene Friends platforms [23]. Open Targets provides an API to cross reference annotations and relationships on diseases, genes, and drugs. Companion genes were pulled from the SIGnaling Network Open Resource (Signor) database due to the manual curation of gene interactions [24].

Therapeutic Drug Data

Therapeutic drug data was obtained from various sources including Drugbank, Finan, et al. ("druggable genome"), and the Drug Gene Interaction Database [25]. Druggable genome data were obtained from the supplementary materials in Finan, et al. in CSV format [26]. The obtained data provided insight on 3000+ potential gene targets with evidence for drug targets or potential targets. DGIdb drug data (accessed January 2023) were downloaded from the DGIdc online database as files consisting of known gene and drug interactions as well as details such as interaction types, drug categories, etc.

Pre-processing

All data pre-processing was carried out using custom Python scripts for data that was not obtained via the Yang Lab or was missing information such as gene symbols (see URLs). Preprocessing included gene annotation, BESD format preparation and conversion, and/or calculation of necessary measures such as beta values.

Gene annotation was performed as necessary if no gene symbol was provided in the original data source. Annotation conducted using both the Python biomart package and pyensembl [27,28] using ENSG IDs. mQTL gene annotation was conducted by obtaining Illumina 450k chip probe data using IlluminaHumanMethylation450kanno.ilmn12.hg19, an R package available through Bioconductor [29].

Data sources that were not obtained in SMR ready format (BESD file format) such as pQTL and multiancestry eQTL data were formatted into BESD format using the flist method outlined by the Yang Lab. BESD format is a format used by the Yang lab to store x-QTL summary data that consists of three files - .esi, .epi, .besd. More information on the format and how to process data into BESD format can be found on the SMR Yang Lab website (https://yanglab.westlake.edu.cn/software/smr/#DataManagement).
Our multi-ancestry eQTL data did not contain allele frequency, beta, or standard error values. Allele frequencies were obtained using the 1000 Genomes reference panel. We then derived beta and standard error values using each eQTL's random effects model z-score, allele frequency and total number of samples from the original study (n = 2,119).

Summary data-based Mendelian Randomization

Summary data-based Mendelian Randomization (SMR) is a MR computational tool that uses summary-level data to test if an exposure variable (i.e., gene expression) and outcome (i.e., trait) are causally associated because of a shared causal variant (i.e. instrumental variable)[30]. A key feature incorporated into the SMR method is the ability to discern potentially causal variants from those in linkage disequilibrium with the functional variant. In order to distinguish pleiotropy from linkage, the heterogeneity in dependent instruments (HEIDI) method was implemented using the default version which uses the top 20 SNPs at a locus. Linkage disequilibrium reference data was obtained from 1000 Genomes phase 3 reference panel [31]. Additionally, in order to increase statistical power we applied the SMR-multiple exposures (SMR-multi) method, an extension of the SMR method, which allows for simultaneous testing of multiple traits or exposures on a single outcome by using a Bayesian framework to estimate the effects of multiple traits or exposures while accounting for the correlation between them. SMR and HEIDI analysis were conducted using the SMR software established and maintained by the Yang Lab using all default parameters including those previously detailed [30,32].

The post SMR-processing first consisted of removal of potential associations with no available gene annotations, associations annotated to "novel genes", associations with genes in the major histocompatibility complex (MHC), or associations not annotated as protein-coding genes. In order to obtain significance we first implemented a

significance threshold of $p_{SMR_multi} < 2.95E-06$ (0.05/16,875). We chose to correct for 16,875 protein coding genes given that is the maximum number of tested genes across all NDD and omic pairs. We then filtered based on the presence of pleiotropy via the computed HEIDI score ($p_{HEIDI} > 0.01$ for inclusion in this study) using a less conservative threshold than the threshold originally used by the creators of the package [30]. SNPs were then split on their associated genes status as a therapeutic target or as a non-therapeutic target. After initial processing, analyses were conducted as demonstrated in our workflow diagram **(Figure 1a)** and explained further in our gene nomination workflow below. In total, when building this reference resource for the community, we tested a total of 186 omic-tissue pairs across six NDDs (**Table 1**).

Disease	Total Genes (Unique)	Liver Genes	Total eQTL Genes (non multi ancestry)	Replicated in Multi Ancestry	Total Therapeutic Genes	% Therapeutic	Total Non-therapeutic Genes	% Non-therapeutic
All Tested Genes (Protein Coding)								
AD	16833	1597	15112	8404	3,562	21.2%	13,271	78.8%
ALS	16875	1610	15163	8408	3,565	21.1%	13,310	78.9%
FTLD	16788	1537	15038	8394	3,551	21.2%	13,237	78.8%
LBD	16797	1540	15069	8388	3,554	21.2%	13,243	78.8%
PD	16872	1596	15159	8407	3,566	21.1%	13,306	78.9%
PSP	16042	1033	13839	8073	3,420	21.3%	12,622	78.7%
Significance p_SMR_multi < 0.05 & p_HEIDI > 0.01								
AD	4948	175	3189	2079	1,142	23.1%	3,806	76.9%
ALS	3188	83	1857	1260	715	22.4%	2,473	77.6%
FTLD	2318	78	1243	810	542	23.4%	1,776	76.6%
LBD	2530	82	1384	900	580	22.9%	1,950	77.1%
PD	3592	108	2161	1434	811	22.6%	2,781	77.4%
PSP	2275	30	1270	842	574	25.2%	1,701	74.8%
Significance p_SMR_multi < 2.95E-06 (testing all protein coding genes) & p_HEIDI > 0.01								
AD	159	4	97	44	31	19.5%	128	80.5%
ALS	4	0	3	3	0	0.0%	4	100.0%
FTLD	0	0	0	0	0	0.0%	0	0.0%
LBD	6	0	2	0	1	16.7%	5	83.3%
PD	71	5	56	26	15	21.1%	56	78.9%
PSP	14	1	8	5	2	14.3%	12	85.7%

Significance p_SMR_multi < 1.58E-08 (testing all protein coding genes across all omics) & p_HEIDI > 0.01								
AD	60	1	39	19	14	23.3%	46	76.7%
ALS	1	0	0	0	0	0.0%	1	100.0%
FTLD	0	0	0	0	0	0.0%	0	0.0%
LBD	2	0	1	0	1	50.0%	1	50.0%
PD	42	2	36	14	8	19.0%	34	81.0%
PSP	13	0	8	5	2	15.4%	11	84.6%

Table 1: Summary of SMR data mining across diseases | AD = Alzheimer's Disease, ALS = Amyotrophic Lateral Sclerosis, FTLD = Frontotemporal Dementia Lobar Degeneration, LBD = Lewy Body Dementia, PD = Parkinson's Disease, PSP = Progressive Supranuclear Palsy

Gene Nomination and Drug Target Identification

Gene nomination was conducted by focusing on identifying shared significant genetic targets between the different NDDs. Gene nomination was then conducted using a system consisting of the three categories previously described in the introduction. Therapeutic targets were initially nominated using data sources from the two previously outlined database resources. Further target curation was conducted using Open Targets to verify if any approved indications included an NDD thus allowing us to classify drugs into either the *novel* or *known* tiers. Identified network companion genes upstream and downstream of the initial target identified were further categorized into groups based on therapeutic status and approved use in treating any NDD. Our *known* and *difficult* tiers were further investigated using gene co-expression networks via Open Targets interaction annotations through the Signor database. Using interaction we identified companion genes (genes that are manually annotated for their causal relationships with the gene of interest by Signor) for our nominated genes. We additionally searched known therapeutics that target any identified companion genes to potentially identify proxy gene targets thus expanding the net for drug discovery and repurposing. We implemented custom python scripts in order to query Open Targets' API to extract relevant annotations for this workflow.

Results

Overview

We identified 540 candidate gene-level SMR associations (159 unique gene targets) across six NDDs and 186 tissue-omic pairings with a stringent disease-level multiple test correction ($p_{SMR_multi} < 2.95E-06$ and $p_{HEIDI} > 0.01$; **Supplementary Table S1**). On a per disease basis we identified 317 significant associations and 116 unique gene targets for AD, 4 significant associations and 3 unique gene targets for ALS, no associations or gene targets for FTLD, 13 significant associations and 5 unique genes for LBD, 184 significant associations and 46 unique gene

targets for PD, and 22 significant associations and 9 unique gene targets for PSP. We were unable to identify significant associations in any pQTL or caQTL omics as well as in ALS or FTLD at our corrected p value threshold.

SMR analysis identifies 15 common genes significant across multiple NDD

Using SMR, we identified 15 unique genes across 182 associations to be significant in two or more NDDs at a stringent significance threshold ($p_{SMR_multi} < 2.95E-06$ and $p_{HEIDI} > 0.01$, **Table 2**, **Supplementary Table S2**, **Supplementary Table S3**). Of the identified genes, five genes, *MAPT, CRHR1, KANSL1, ARL17A,* and *ARHGAP27*, were found to be significant across 97 tested associations and three NDDs (AD, PD, and PSP). *MAPT* and *CRHR1* were found to be largely significant in mQTL omics with *MAPT* significant in whole blood and brain mQTL data for all previously mentioned NDDs and *CRHR1* found to be significant in whole blood mQTL data for all three NDDs (**Supplementary Tables S2 - S4**). Additionally, *MAPT* and *CRHR1* are considered druggable in multiple drug data sources while the remaining three genes are not considered druggable as of the writing of this manuscript. All aforementioned genes, except for *ARL17A*, had multiple significant associations in both brain and blood mQTL tissues. *ARHGAP27* and *KANSL1* had significant associations replicated in our multi-ancestry cohort (African, American, East Asian, European, and South Asian ancestries) with both genes showing an associated decreased expression when disease risk increased by a standard deviation from the population mean genetic risk for that specific disease suggesting generalizability across populations.

Gene	**Diseases**	**Omics**
ARL17B	AD, PD	Cerebellum eQTL, Cortex eQTL, Spinalcord eQTL
KAT8	AD, PD	Cerebellum eQTL, Whole Brain meta-analysis mQTL, Cerebellar Hemisphere eQTL, Cortex eQTL, Tibial Nerve eQTL, Skeletal Muscle eQTL, Hypothalamus eQTL, Whole Brain eQTL, Cerebellum eQTL, Spinalcord eQTL
LRRC37A2	AD, PD	Hippocampus eQTL, Cortex eQTL, Frontal Cortex BA9 eQTL, Prefrontal Cortex eQTL, Caudate Basal Ganglia eQTL, Skeletal Muscle eQTL, Multi Ancestry, Whole Brain Meta-analysis eQTL, Hypothalamus eQTL, Liver eQTL, Anterior Cingulate Cortex BA24 eQTL, Putamen Basal Ganglia eQTL, Amygdala eQTL, Whole Brain eQTL, Cerebellum eQTL, Nucleus Accumbens eQTL, Basal Ganglia eQTL, Spinalcord eQTL, Hippocampus eQTL, Substantia nigra eQTL
KANSL1	AD, PD, PSP	Whole Brain meta-analysis mQTL, Whole Blood mQTL, Cortex eQTL, Multi Ancestry Whole Brain Meta-analysis eQTL, Spinalcord eQTL, Anterior Cingulate Cortex BA24 eQTL
ARL17A	AD, PD, PSP	Spinalcord eQTL, Amygdala eQTL, Multi Ancestry Whole Brain Meta-analysis eQTL, Hypothalamus eQTL, Hippocampus eQTL, Cerebellar Hemisphere eQTL, Cortex eQTL, Caudate Basal Ganglia eQTL, Anterior Cingulate Cortex BA24 eQTL, Putamen Basal Ganglia eQTL, Cerebellum eQTL, Nucleus Accumbens Basal Ganglia
PRSS36	AD, PD	Whole Brain meta-analysis mQTL, Cortex eQTL, Cerebellar Hemisphere eQTL, Multi Ancestry Whole Brain Meta-analysis eQTL, Whole Brain eQTL
MAPT	AD, PD, PSP	Whole Brain meta-analysis mQTL, Whole Blood mQTL
IDUA	LBD, PD	Whole Brain meta-analysis mQTL, Whole Blood mQTL, Whole Blood eQTL(eQTLgen)

TMEM175	LBD, PD	Whole Blood mQTL
ARHGAP27	AD, PD, PSP	Whole Blood mQTL, Whole Blood eQTL (eQTLgen), Multi Ancestry Whole Brain Meta-analysis eQTL, Caudate Basal Ganglia eQTL, Nucleus Accumbens Basal Ganglia
CRHR1	AD, PD, PSP	Whole Brain meta-analysis mQTL, Whole Blood mQTL, Cortex eQTL, Skeletal Muscle eQTL
FMNL1	AD, PSP	Multi Ancestry Whole Brain Meta-analysis eQTL, Whole Blood mQTL
PLEKHM1	PD, PSP	Cortex eQTL, Frontal Cortex BA9 eQTL, Prefrontal Cortex eQTL, Caudate Basal Ganglia eQTL, Skeletal Muscle eQTL, Anterior Cingulate Cortex BA24 eQTL, Putamen Basal Ganglia eQTL, Whole Brain eQTL
WNT3	AD, PD	Cortex eQTL metaBrain, Skeletal Muscle eQTL, Tibial Nerve eQTL
SPPL2C	AD, PD	Cerebellum eQTL, Prefrontal Cortex eQTL

Table 2: Candidate genes for multiple neurodegenerative diseases. This table shows genes with functional inferences passing multiple test correction at a multi-SNP SMRP < 2.95E-06 for multiple neurodegenerative diseases. We provide details for all the omics and diseases in which a given gene has significant associations.

We identified 10 unique genes across 85 significant associations in any two NDDs (p_{SMR_multi} < 2.95E-06 and p_{HEIDI} > 0.01). One of the identified genes is considered therapeutic and the remaining nine are non-therapeutic. *KAT8*, *ARL17B, PRSS36, LRRC37A2, WNT3,* and *SPPL2C* were all found to be significant in both AD and PD, *IDUA* and *TMEM175* were found to be significant in LBD and PD, *PLEKHM1* was significant in *PD* and *PSP*, and *FMNL1* was significant in AD and PSP (**Supplementary Tables, S2, S3, S5**). *KAT8,* the only therapeutic gene*,* showed significant associations in both mQTL tissues where it had an associated increased expression compared to the other omic tissues. Of the remaining nine genes, *IDUA, FMNL1, PRSS36,* and *TMEM175* had significant associations in mQTL sources. Additionally, *FMNL1, LRRC37A2,* and *PRSS36* had significant associations replicated in our multi ancestry cohort (p_{SMR_multi} < 2.95E-06; **Supplementary Table S6**).

Drug Target Discovery using significant genes identifies 41 novel gene targets for follow-up study.

Using the approach previously outlined in our introduction and methods for drug target gene nomination, we categorized 159 gene hits into one of three tiers (**Table 3**). In our first tier, *novel* genes, we nominated 41 gene targets which are listed in **Supplementary Table S7**. Current literature, knowledge base, and drug databases do not identify any therapeutics targeting *novel* genes that are FDA approved for treatment in any NDDs and in druggable regions of the genome able to be targeted by common molecular methods. Our second tier, *known* genes, had three gene targets identified - *MAPT, KCNN4,* and *ADORA2B,* indicating that these genes have at least one therapeutic that has FDA approved use for treatment on an NDD (**Supplementary Table S8**). These gene targets are targeted by four therapeutics - Apomorphine, Carbidopa, Istradefylline, and Riluzole. The two diseases targeted by the aforementioned therapeutics are ALS and PD. In our last and largest tier, *difficult genes*, we identified 115 gene targets with no currently known therapeutics that target these genes. A total of 52 of the identified *difficult* genes

were associated with at least two significant associations, with *LRRC37A2* having the maximum number of significant associations at 25 genes across AD and PD.

Tier	Requirements	# of Genes	Genes
Novel	- Druggable - Not approved for use in treating NDD	41	ADAM10, SNCA, EGFR, POU5F1, STK39, INPP5D, CRHR1, APH1B, MINK1, CLU, CR1, ACE, CD38, RABEP1, ERCC2, KAT8, ITGAX, GAK, STX4, EPHB4, EPHA1, GPNMB, STAG3, CHRNE, NDUFS2, FCER1G, VKORC1, DNTT, CKM, HSD3B7, BST1, STX1B, PSMC3, CDSN, MICB, MS4A2, PSORS1C1, EPHX2, SLC44A4, MAT1A, FBXL19
Known	- Druggable - Approved for use in treating NDD	3	MAPT, KCNN4, ADORA2B
Difficult	- No known druggability	115	TRIM27, PPP4C, SPI1, EFNA3, KIF1C, WNT3, CD2AP, CCNE2, KCTD13, C9orf72, SRCAP, CELF1, HIP1R, GRN, APOC2, ARHGAP27, MEPCE, LRRFIP2, COPS6, GIGYF1, BCKDK, POLR2E, EFNA4, DYDC1, ATF6B, LLGL1, MTMR2, GPC2, LRRC37A, ARL17B, INO80E, SNX31, CEACAM19, DGKQ, NUP42, LRRC37A2, KANSL1, ARL17A, ANXA11, TSPAN14, CASTOR3, ZNF232, ZNF45, TSBP1, TREM2, PRSS36, IDUA, CCDC158, CCDC189, ZSWIM7, PLEKHM1, STH, PVRIG, YPEL3, MMRN1, SPPL2C, SCIMP, PILRB, PILRA, LACTB, FMNL1, APOC4, ZNF646, CPSF3, ZSCAN9, ZKSCAN3, TREML2, EPDR1, UFSP1, FAM131B, TAS2R60, USP6NL, MS4A4A, CASS4, G2E3, SCFD1, PCGF3,SETD1A,DCAKD, ZNF668, AGFG2, TMEM175, TOMM40, TRIM40, WDR81, TMEM106B, FNBP4, SHROOM3, CYP21A2, REXO1, TNXB, MS4A3, AIF1, RAB8B, ZFP57, FAM200B, BTNL2, IGSF9B, HS3ST1, ZNF311, NDUFAF6, TMEM163, APOC1, C17orf107, EXOC3L2, DYDC2, DOC2A, ACMSD, TRIM31, PRDM7, TRIM10, ZAN, MS4A6A, CPLX1, SFTA2

Table 3: Therapeutic Classification Scheme by Tier. Table providing information on the three classifications tiers in our therapeutic classification scheme including requirements for each tier. The number of genes in each tier and which genes are in each tier are provided.

Network analysis provides insight into druggable companion genes to non-druggable genes of interest.

We further implemented a gene network analysis for our *novel* and *difficult* tier candidates to identify potential proxy gene targets within each nominated genes Signor curated network. In the *novel* tier gene we identified 87 companion genes of which 58 are considered potentially druggable (**Supplementary Table S8**). Of the 58 druggable companion genes, 30 were found to be targeted by a *known* drug and a further five are targeted by therapeutics approved for treatment of AD. The five companion genes with AD targeted therapeutics are *NCSTN, MAPK14, PSEN1, PSEN2,* and *PSENEN*. Genes *NCSTN, PSEN1, PSEN2,* and *PSENEN* are all targeted by Tarenflurbil, Semagacestat, and Avagacestat while *MAPK14* is targeted by Neflamapimod, an oral p38 alpha kinase inhibitor FDA approved for use in the treatment of AD and LBD. Further analysis of *difficult* genes co-expression networks

identified 27 genes with 65 curated companion genes (**Supplementary Table S10**, **Supplementary Table S11**). Of the 65 identified companion genes, 34 were found to be druggable with 18 having *known* drugs. *MAPK14* was the only companion gene to have a therapeutic approved to treat an NDD. *MAPK14* was identified as a companion gene to the *difficult* gene *TRIM27* and is targeted by Neflamapimod.

Multi-ancestry analyses reveal opposing gene expression patterns in significant disease risk loci between Non-European and European ancestries.

Multi ancestry eQTL data was compared against all other eQTL data sources in order to replicate significant hits ($P_{SMR_multi} < 2.95E\text{-}06$ & $P_{HEIDI} > 0.01$). Overall, our validation analysis included 8,536 multi ancestry eQTL hits and 15,183 non multi ancestry eQTL hits. In total, there were 11 significant multi ancestry SNPs and 99 significant SNPS in the non multi ancestry eQTL data; we identified nine replicated significant hits, *ARHGAP27*, *ARL17A*, *GPNMB*, *KANSL1*, *LRRC37A2*, *PILRA*, *PILRB*, *PRSS36*, and *ZNF232* (**Supplementary Table S6**). Additionally, we split the nine replicated hits by their known druggable status and non-druggable status. We identified one replicated significant druggable hit, *GPNMB*, and the remaining 8 replicated hits were non-druggable.

Discussion

As the global population continues to age, the threat posed by NDD presents a behemoth and multifaceted challenge. Our research aims to address the challenge of treating NDD by identifying therapeutic targets anchored in genetic data - a proven strategy in therapeutic development. Implementation of this strategy has been impeded by the small sample size and dispersed nature of the genetic and disease related data, such as proteomics and transcriptomics, let alone in a harmonized fashion. Here we attempted to address this need by creating and implementing an open source framework to identify druggable targets across varied NDDs.

In our targeted analyses, we were unable to identify any significant and potentially causal genes present across all six NDDs. While NDDs share prominent hallmarks, such as cell death, inflammation, and pathological protein aggregation, the role that each hallmark and its associated biological processes plays in the pathogenesis of each NDD differs, creating a spectrum [33,34]. We identified *MAPT*, *CRHR1*, *KANSL1*, *ARL17A*, and *ARHGAP27* to be significant in multiple different omics for AD, PD, and PSP (**Supplementary Tables S2-S4**). *MAPT* was found to have significant associations with primarily increased expression for AD, PD, and PSP across eQTL and mQTL omic data, as supported by previous research [35-37]. The *MAPT* locus, 17q21, contains genes *CRHR1*, *KANSL1*, *ARL17A*, and *ARHGAP27*, and mutations in this locus have been previously associated with both PD and PSP [38]. Previous evidence of significant association in AD is more fragmented and sparse. Evidence of AD association includes genes *KANSL1* and *MAPT* within the 17q21 locus [39]. *ARL17A* has been reported to harbor eQTL SNPs implicated in both brain and blood tissues in relation to AD [40]. *CRHR1*'s role in stress response has been

hypothesized to exacerbate AD pathologies given its abundance in the brain including areas implicated in learning and memory [41]. Lastly, evidence of *ARHGAP27*'s significance in AD includes associations between complex traits such as cognitive functioning, reaction time, and cortical structure phenotypes [42,43].

A deep dive into *KANSL1* highlights its role in autophagy pathways. *KANSL1* is a core member of the nonspecific lethal (NSL) complex that binds to MOF (also known as *KAT8*) which is necessary for the acetylation of histone H4 lysine 16 acetylation (H4K16ac)[44,45]. Some studies have associated elevated expression levels of *KANSL1* with over promoted autophagic activity resulting in cell death and cytotoxicity from autophagosome accumulation;, although further research is required to understand this mechanism [46]. Additional research into the role of autophagy and lysosomal pathways in NDDs have indicated that altered autophagy function results in the inability to clear out protein aggregates resulting in cell death and potentially contributing to disease pathogenesis and neurodegeneration. [45,47-49]. Our results are consistent with previous research linking increased expression of *KANSL1* with neurodegenerative effects. When assessing associations with AD, PD, and PSP, *KANSL1* is associated with an increased expression in brain mQTLs, three different brain eQTLs (psychEncode, multi ancestry, and anterior cingulate cortex), and spinal cord eQTLs. The consistent significance of *KANSL1* and the majority of our gene hits in mQTL omics highlights the influence of DNA methylation for NDD pathogenesis and progression.

We identified 10 genes as significant in two diseases. The nominated genes do not share any explicit relationships, but are common in their importance for varying biological processes and cellular functions such as cell proliferation and differentiation, degradation of transmembrane proteins, calcium homeostasis, and autophagy regulation [50-54]. Six of our nominated genes, *ARL17B, KAT8, LRRC37A2, PRSS36, SPPL2C, and WNT3*, are associated with both AD and PD. Given the significantly larger sample sizes and increased power of the two diseases in GWAS summary statistics, we did not find this unexpected. LBD and PD share two genes, *IDUA* and *TMEM175,* while AD and PSP share *FMNL1* and PD and PSP share *PLEKHM1* (**Supplementary Tables S2, S3, S5**). In general, the bulk of the gene hits were found to be significant in mQTL data for both brain and blood tissues ($n_{\text{whole brain}}= 4$; $n_{\text{whole blood}}= 4$) followed by cortex eQTLs ($n_{\text{cortex metaBrain}} = 6$, $n_{\text{cortex GTEx}} = 3$, $n_{\text{Frontal Cortex BA9}} = 2$, $n_{\text{prefrontal cortex}} = 3$).

The only gene found in two diseases, AD and PD, that could be targeted therapeutically was *KAT8,* which we previously mentioned in the context of the *KANSL1* gene. In literature, *KAT8* (Lysine Acetyltransferase 8) is identified as a protein-coding gene that plays a vital role in the NSL complex for acetylation of H4K16ac [49]. Scientific observation has identified the consequences of autophagic dysfunction in NDDs to include impaired neuronal function, neuronal death, and neuron loss. In opposition to the expression pattern of *KANSL1*, decreased expression of *KAT8* is associated with deacetylation of H4K16ac in AD patients, while an overexpression of the two has been linked to increased expression levels of neuroprotective soluble amyloid precursor protein (sAPP)α and β-secretase (BACE)2 [55]. In our results, we found blood mQTLs for AD and brain mQTLs for AD and PD to be associated with increased expression of *KAT8*; this is in contrast to gene expression in some of the same tissues, such as blood and brain mQTLs, for *KANSL1*. The associated increased expression of *KAT8* in our results suggest

that an increase in expression may be correlated with excess autophagy resulting in cell death, which is a hallmark symptom of all three NDDs (AD, PD, and LBD; 56,57). There currently does not exist any FDA approved therapeutics that target *KAT8* in NDDs. However, compound MG149, a histone acetyltransferase inhibitor, has been found to reduce proinflammatory genes via inhibition of MYST type histone acetyltransferase KAT8 [58[58]]. MG149 has also been found to be effective in restoring impaired autophagic flux via the inhibition of histone acetylation of H4K16ac in cases of ischemic stroke and inflammatory diseases [48,59]. Further research into the application of MG149 could result in a novel treatment targeting the characteristic accumulation of toxic proteins in NDDs.

FTLD was the only tested disease that did not have any suggestive targets at our test correction threshold. This may be due to the fact that the FTLD GWAS had the smallest sample size out of all the diseases tested, and results will likely improve as larger FTLD GWAS are conducted. As there were no significant results for FTLD after correction, we decided to investigate potential pleiotropic relationships between FTLD and the other NDDs. To do this, we looked for FTLD associations at a less-stringent P value threshold ($p_{SMR_multi} < 0.05$) only in the 254 unique candidate genes passing our original threshold of $p_{SMR_multi} < 2.95E-06$, a process detailed by Baird et al. [60] This resulted in 124 FTLD hits made up of 31 unique genes that have a potential pleiotropic relationship between FTLD and another NDD. Of those 31, 12 were classified as druggable through our sources (*STX4*, *STX1B*, *VKORC1*, *POU5F1*, *HSD3B7*, *PSORS1C1*, *SLC44A4*, *CD38*, *EPHX2*, *FBXL19*, *CLU*, *CDSN*). All 12 fall into the novel tier of drug targets, representing potential avenues for drug repurposing for FTLD.

Our creation of a drug target classification scheme is an attempt to inform drug discovery and repurposing from genes considered significant with evidence of causative roles in NDDs. Further inspection of our 41 *novel* genes provides multiple insights into the genes that compose the tier. The majority of genes that compose our novel tier have therapeutics used in the treatment of multiple types of cancers and tumors. Fourteen of our novel genes have therapeutics approved for use in the treatment of cancer (MONDO_0004992). Other common approved indications for therapeutics that target our novel genes include, but are not limited to, neoplasm (EFO_0000616), hypertension (EFO_0000537), and cardiovascular disease (EFO_0000319). *GPNMB,* which is of particular interest due to support for its role in PD*,* falls into this grouping of 14 genes. Similar to its role in cancer and tumor growth, our results highlight *GPNMB*'s pattern of increased expression as shown in brain related PD eQTLs. We were able to find replication of increased GPNMB expression in brain related tissues in Li et al, Ortiz et al, and Nalls et all [61-63]. Glembatumumab Vedotin is one of the therapeutics that targets *GPNMB* where its primary mechanism of action (MOA) is Tubulin inhibition [23]. Consequently, Glembatumumab Vedotin's inhibitory MOA could be repurposed for use in PD treatment for suppression of inflammation given the recognized role of inflammatory response/neuroinflammation in PD onset and progression [64,65]. However, any treatment developed targeting *GPNMB*, would most likely be limited in treating people of European ancestries due to the gene's importance and role compared to non-European ancestries - further increasing inequality.

Our largest and most uncertain classification tier contains 121 *difficult* genes. Despite not having any currently known therapeutics, this classification tier could lead to the development of NDD targeted therapeutics or the repurposing of existing ones. Our approach for these genes focused on analyzing well curated networks centered on each *difficult* gene in order to identify any partner genes with existing therapeutic drugs. This approach provides us context into any biological pathways and processes that may be affected by a targeted treatment which could help eliminate the time and resources spent on developing and researching ineffective therapies.

The smallest tier, *known* genes, is composed of the three genes targeted by NDD targeted therapeutics. Apomorphine, Carbidopa, and Istradefylline are indicated for use in treatment of PD. Riluzole is indicated for the treatment of ALS but has undergone phase 2 clinical trials for use in treatment of AD. The results in clinical trials for use of Riluzole in AD treatment were promising with cerebral glucose metabolism, an AD biomarker, preserved in patients receiving riluzole compared to those in the placebo group [66]. The researchers conducting the study suggested a more powerful and longer study but no follow up studies have yet been initiated. Our results support the continued follow up of Riluzole clinical trials.

Genes such as *GPNMB* had different expression patterns in European and non European ancestries. For example, *GPNMB* had decreased associated expression in multi ancestry eQTLs but an increased associated expression in all other tested eQTLs. Previous research in certain Asian populations has found no significant association between *GPNMB* and PD [67,68]. Rizig and colleagues, conducting the largest PD GWAS in the African and African admixed populations in ~200,000 individuals, of which 1,488 are cases, report the following per SNP in *GPNMB*: rs858275, P=0.1250, beta=-0.0824, indicating no association in African/African admixed ancestries. Our multi ancestry data reports the same direction of expression in GPNMB SNP rs858275, P=1.080397e-08, beta=-0.107745 in PD. Interestingly, the reported direction of expression in our multi ancestry data and Rizig and colleagues data is in contrast to the direction of expression reported for European ancestries, in addition to indicating no significant associations (**Supplementary Table S12, under preparation**).

The limitations we encountered in our research included limited GWAS data for diseases excluding AD and PD, limited non-eQTL omic data, as well as limited multi ancestry omic data and reference panels. In general, the availability of publicly and freely available omic data is consistently increasing. As new data is published we intend to conduct updates and incorporate new omic types into our analysis such as more pQTL, single cell QTLs, and splicing quantitative trait loci (sQTLs). The incorporation of additional multi omic data should provide new and novel insights into the complex underpinnings of NDD.

The limitation we feel presents the most barriers is that of limited multi ancestry data. The state of diversity in the NDD research space has historically been eurocentric which remains the case in this study due to the limited availability of genetic data from non-European participants. One of the distinguishing aspects of this study is the inclusion of multi ancestry eQTL data in the search for generalizable drug targets. This is particularly important in

an era where precision medicine and machine learning can introduce inherent bias when the only reference data is from European ancestry populations. We identified common hits which are consistent with current understanding that there are NDD risk loci that are shared across genetic ancestries while providing insight on which gene loci and differences in expression may play a role in NDD development and treatment in non-Europeans. It is worth noting that while replication was limited at our stringent significance threshold we were able to make some interesting observations. While we made attempts to include a limited set of multi ancestry in the future we would like to be able to include more multi ancestry disease GWAS and omic data in order to make more meaningful insights. We look forward to the increasing availability of non-European data with the creation of data sources such All of Us, an NIH research program focusing on inclusion of health data of marginalized populations in the United States [69].

This report is a description of the foundation for a community driven resource to identify and investigate future genetically derived drug targets in an open source context. Ultimately, we are working on creating a network tool that incorporates multi-omic data, disease GWAS summary statistics, drug data, and other relevant data types to ease research such as this study; eliminating barriers to drug discovery and drug repurposing and potentially enabling precision medicine in the NDD space. Using multi omics integration methods, deep learning techniques, and most importantly, community input to better parse and interpret the data presented by the platform we aim to make our community resource a robust tool for NDD research.

References

1. van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *N. Engl. J. Med.* **388**, 9–21 (2023).
2. Morant, A. V., Jagalski, V. & Vestergaard, H. T. Labeling of Disease-Modifying Therapies for Neurodegenerative Disorders. *Front. Med.* **6**, 223 (2019).
3. Dementia. https://www.who.int/news-room/fact-sheets/detail/dementia.
4. King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* **15**, e1008489 (2019).
5. Paulsen, J. S. *et al.* A review of quality of life after predictive testing for and earlier identification of neurodegenerative diseases. *Prog. Neurobiol.* **110**, 2–28 (2013).
6. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).

7. Chia, R. *et al.* Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat. Genet.* **53**, 294–303 (2021).

8. Höglinger, G. U. *et al.* Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet.* **43**, 699–705 (2011).

9. Nalls, M. A. *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.* **18**, 1091–1102 (2019).

10. Nicolas, A. *et al.* Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron* **97**, 1268–1283.e6 (2018).

11. Pottier, C. *et al.* Genome-wide analyses as part of the international FTLD-TDP whole-genome sequencing consortium reveals novel disease risk factors and increases support for immune dysfunction in FTLD. *Acta Neuropathol.* **137**, 879–899 (2019).

12. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

13. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).

14. de Klein, N. *et al.* Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat. Genet.* (2023) doi:10.1038/s41588-023-01300-6.

15. Zeng, B. *et al.* Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat. Genet.* **54**, 161–169 (2022).

16. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

17. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, (2018).

18. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).

19. Vaja, R. & Rana, M. Drugs and the liver. *Anaesth. Intensive Care Med.* **21**, 517–523 (2020).

20. McRae, A. F. *et al.* Identification of 55,000 Replicated DNA Methylation QTL. *Sci. Rep.* **8**, 17605 (2018).

21. Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, 3121 (2018).

22. Yang, C. *et al.* Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat. Neurosci.* **24**, 1302–1312 (2021).

23. Ochoa, D. *et al.* The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2023).

24. Lo Surdo, P. *et al.* SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update. *Nucleic Acids Res.* **51**, D631–D637 (2023).

25. Freshour, S. L. *et al.* Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res.* **49**, D1144–D1151 (2021).

26. Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, (2017).

27. GitHub - openvax/pyensembl: Python interface to access reference genome features (such as genes, transcripts, and exons) from Ensembl. *GitHub* https://github.com/openvax/pyensembl.

28. GitHub - sebriois/biomart: Python biomart API. *GitHub* https://github.com/sebriois/biomart.

29. IlluminaHumanMethylation450kanno.ilmn12.hg19. *Bioconductor* http://bioconductor.org/packages/IlluminaHumanMethylation450kanno.ilmn12.hg19/.

30. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

31. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

32. Wu, Y. *et al.* Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* **9**, 918 (2018).

33. Seo, J. & Park, M. Molecular crosstalk between cancer and neurodegenerative diseases. *Cell. Mol.*

Life Sci. **77**, 2659–2680 (2019).

34. Wilson, D. M., 3rd *et al.* Hallmarks of neurodegenerative diseases. *Cell* **186**, 693–714 (2023).
35. Zhou, F. & Wang, D. The associations between the MAPT polymorphisms and Alzheimer's disease risk: a meta-analysis. *Oncotarget* **8**, 43506–43520 (2017).
36. Tobin, J. E. *et al.* Haplotypes and gene expression implicate the MAPT region for Parkinson disease: the GenePD Study. *Neurology* **71**, 28–34 (2008).
37. Majounie, E. *et al.* Variation in tau isoform expression in different brain regions and disease states. *Neurobiol. Aging* **34**, 1922.e7–1922.e12 (2013).
38. Lin, M. K. & Farrer, M. J. Genetics and genomics of Parkinson's disease. *Genome Med.* **6**, 48 (2014).
39. Jun, G. *et al.* A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol. Psychiatry* **21**, 108–117 (2016).
40. Patel, D. *et al.* Cell-type-specific expression quantitative trait loci associated with Alzheimer disease in blood and brain tissue. *Transl. Psychiatry* **11**, 250 (2021).
41. Futch, H. S., Croft, C. L., Truong, V. Q., Krause, E. G. & Golde, T. E. Targeting psychologic stress signaling pathways in Alzheimer's disease. *Mol. Neurodegener.* **12**, 49 (2017).
42. Zhao, B. *et al.* Transcriptome-wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits. *Nat. Commun.* **12**, 2878 (2021).
43. Madrid, L., Labrador, S. C., González-Pérez, A., Sáez, M. E. & The Alzheimer's Disease Neuroimaging Initiative Adni. Integrated Genomic, Transcriptomic and Proteomic Analysis for Identifying Markers of Alzheimer's Disease. *Diagnostics (Basel)* **11**, (2021).
44. Sheikh, B. N., Guhathakurta, S. & Akhtar, A. The non-specific lethal (NSL) complex at the crossroads of transcriptional control and cellular homeostasis. *EMBO Rep.* **20**, e47630 (2019).
45. Linda, K. *et al.* Imbalanced autophagy causes synaptic deficits in a human model for neurodevelopmental disorders. *Autophagy* **18**, 423–442 (2022).
46. Button, R. W., Roberts, S. L., Willis, T. L., Hanemann, C. O. & Luo, S. Accumulation of

autophagosomes confers cytotoxicity. *J. Biol. Chem.* **292**, 13599–13614 (2017).

47. Son, S. M., Park, S. J., Fernandez-Estevez, M. & Rubinsztein, D. C. Autophagy regulation by acetylation-implications for neurodegenerative diseases. *Exp. Mol. Med.* **53**, 30–41 (2021).
48. Lingling, D., Miaomiao, Q., Yili, L., Hongyun, H. & Yihao, D. Attenuation of histone H4 lysine 16 acetylation (H4K16ac) elicits a neuroprotection against ischemic stroke by alleviating the autophagic/lysosomal dysfunction in neurons at the penumbra. *Brain Res. Bull.* **184**, 24–33 (2022).
49. Füllgrabe, J. *et al.* The histone H4 lysine 16 acetyltransferase hMOF regulates the outcome of autophagy. *Nature* **500**, 468–471 (2013).
50. Wiesel-Motiuk, N. & Assaraf, Y. G. The key roles of the lysine acetyltransferases KAT6A and KAT6B in physiology and pathology. *Drug Resist. Updat.* **53**, 100729 (2020).
51. Wang, C., Telpoukhovskaia, M. A., Bahr, B. A., Chen, X. & Gan, L. Endo-lysosomal dysfunction: a converging mechanism in neurodegenerative diseases. *Curr. Opin. Neurobiol.* **48**, 52–58 (2018).
52. Kragh, C. L., Ubhi, K., Wyss-Coray, T. & Masliah, E. Autophagy in dementias. *Brain Pathol.* **22**, 99–109 (2012).
53. McEwan, D. G. *et al.* PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins. *Mol. Cell* **57**, 39–54 (2015).
54. Lichtenthaler, S. F., Lemberg, M. K. & Fluhrer, R. Proteolytic ectodomain shedding of membrane proteins in mammals-hardware, concepts, and recent developments. *EMBO J.* **37**, (2018).
55. Chen, F. *et al.* miR-149-5p inhibition reduces Alzheimer's disease β-amyloid generation in 293/APPsw cells by upregulating H4K16ac via KAT8. *Exp. Ther. Med.* **20**, 88 (2020).
56. Erekat, N. S. Apoptosis and its Role in Parkinson's Disease. in *Parkinson's Disease: Pathogenesis and Clinical Aspects* (eds. Stoker, T. B. & Greenland, J. C.) (Codon Publications).
57. Goel, P. *et al.* Neuronal cell death mechanisms in Alzheimer's disease: An insight. *Front. Mol. Neurosci.* **15**, 937133 (2022).
58. van den Bosch, T. *et al.* A 6-alkylsalicylate histone acetyltransferase inhibitor inhibits histone acetylation and pro-inflammatory gene expression in murine precision-cut lung slices. *Pulm.*

Pharmacol. Ther. **44**, 88–95 (2017).

59. Dekker, F. J., van den Bosch, T. & Martin, N. I. Small molecule inhibitors of histone acetyltransferases and deacetylases are potential drugs for inflammatory diseases. *Drug Discov. Today* **19**, 654–660 (2014).
60. Baird, D. A. *et al.* Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLoS Genet.* **17**, e1009224 (2021).
61. Li, Y. I., Wong, G., Humphrey, J. & Raj, T. Prioritizing Parkinson's disease genes using population-scale transcriptomic data. *Nat. Commun.* **10**, 994 (2019).
62. Diaz-Ortiz, M. E. *et al.* GPNMB confers risk for Parkinson's disease through interaction with α-synuclein. *Science* **377**, eabk0637 (2022).
63. Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
64. Pajares, M., I Rojo, A., Manda, G., Boscá, L. & Cuadrado, A. Inflammation in Parkinson's Disease: Mechanisms and Therapeutic Implications. *Cells* **9**, (2020).
65. Marogianni, C. *et al.* Neurodegeneration and Inflammation-An Interesting Interplay in Parkinson's Disease. *Int. J. Mol. Sci.* **21**, (2020).
66. Matthews, D. C. *et al.* Riluzole, a glutamate modulator, slows cerebral glucose metabolism decline in patients with Alzheimer's disease. *Brain* **144**, 3742–3755 (2021).
67. Wu, H.-C. *et al.* DLG2, but not TMEM229B, GPNMB, and ITGA8 polymorphism, is associated with Parkinson's disease in a Taiwanese population. *Neurobiol. Aging* **64**, 158.e1–158.e6 (2018).
68. Xu, Y. *et al.* No association of GPNMB rs156429 polymorphism with Parkinson's disease, amyotrophic lateral sclerosis and multiple system atrophy in Chinese population. *Neurosci. Lett.* **622**, 113–117 (2016).
69. All of Us Research Hub. https://www.researchallofus.org/.

Author Contributions

M.A.N., D.V., H.L.L., F.F., and H.I. conceived and planned the main conceptual ideas. M.A.N., H.L.L., and F.F supervised the project. D.V., C.X.A, and M.A.N. performed all initial data collection, wrangling and computations. C.X.A. processed and updated all experimental data, performed the analysis, and designed the figures. H.L.L., M.B.M., and M.A.N. aided in interpreting analyses results. C.X.A, M.B.M., and M.A.N. contributed to the drafting of the manuscript and preparation for publication. A.S., K.L., M.J.K., and S.B.C. consulted and provided critical feedback to the manuscript. All authors revised the manuscript.

Acknowledgements and Funding

We would like to thank all of the subjects who donated their time and biological samples to the studies that are reflected in these analyses. This research was supported in part by the Intramural Research Program of the National Institutes of Health (National Institute on Aging and National Institute of Neurological Disorders and Stroke; project numbers: 1ZIAAG00935, 1ZIANS003154, Z01-AG000949-02). This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (http://hpc.nih.gov).

Data Availability

Omic data and GWAS summary statistics which we have expanded on in our Methods section are freely and publicly available. An extensive listing of availability, as well as direct links to the data sources, can be located here (https://nih-card-ndd-smr-home-syboky.streamlit.app/About). All code can be found on the NIH CARD github linked here (https://github.com/NIH-CARD/NDD_SMR). All SMR associations and annotations are available to browse using an interactive web application (https://nih-card-ndd-smr-home-syboky.streamlit.app/).

Additional Information

Competing Interests

CXA, DV, KL, HLL, FF, and MAN declare that they are consultants employed by Data Tecnica International, whose participation in this is part of a consulting agreement between the US National Institutes of Health and said company. MAN also an advisor to Neuron23 Inc and Character Biosciences.