

RESEARCH

Open Access



AlzGPS: a genome-wide positioning systems platform to catalyze multi-omics for Alzheimer's drug discovery

Yadi Zhou^{1†}, Jiansong Fang^{1†}, Lynn M. Bekris^{1,2}, Young Heon Kim¹, Andrew A. Pieper^{3,4,5,6,7,8}, James B. Leverenz⁹, Jeffrey Cummings^{10,11} and Feixiong Cheng^{1,2,12*} 

Abstract

Background: Recent DNA/RNA sequencing and other multi-omics technologies have advanced the understanding of the biology and pathophysiology of AD, yet there is still a lack of disease-modifying treatments for AD. A new approach to integration of the genome, transcriptome, proteome, and human interactome in the drug discovery and development process is essential for this endeavor.

Methods: In this study, we developed AlzGPS (Genome-wide Positioning Systems platform for Alzheimer's Drug Discovery, <https://alzgps.lerner.ccf.org>), a comprehensive systems biology tool to enable searching, visualizing, and analyzing multi-omics, various types of heterogeneous biological networks, and clinical databases for target identification and development of effective prevention and treatment for AD.

Results: Via AlzGPS: (1) we curated more than 100 AD multi-omics data sets capturing DNA, RNA, protein, and small molecule profiles underlying AD pathogenesis (e.g., early vs. late stage and tau or amyloid endophenotype); (2) we constructed endophenotype disease modules by incorporating multi-omics findings and human protein-protein interactome networks; (3) we provided possible treatment information from ~ 3000 FDA approved/investigational drugs for AD using state-of-the-art network proximity analyses; (4) we curated nearly 300 literature references for high-confidence drug candidates; (5) we included information from over 1000 AD clinical trials noting drug's mechanisms-of-action and primary drug targets, and linking them to our integrated multi-omics view for targets and network analysis results for the drugs; (6) we implemented a highly interactive web interface for database browsing and network visualization.

(Continued on next page)

* Correspondence: chengf@ccf.org

[†]Yadi Zhou and Jiansong Fang contributed equally to this work.

¹Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

²Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: Network visualization enabled by AlzGPS includes brain-specific neighborhood networks for genes-of-interest, endophenotype disease module networks for omics-of-interest, and mechanism-of-action networks for drugs targeting disease modules. By virtue of combining systems pharmacology and network-based integrative analysis of multi-omics data, AlzGPS offers actionable systems biology tools for accelerating therapeutic development in AD.

Keywords: Alzheimer's disease, Clinical trial, Drug repurposing, Genomics, Mechanism-of-action, Multi-omics, Network medicine, Systems pharmacology

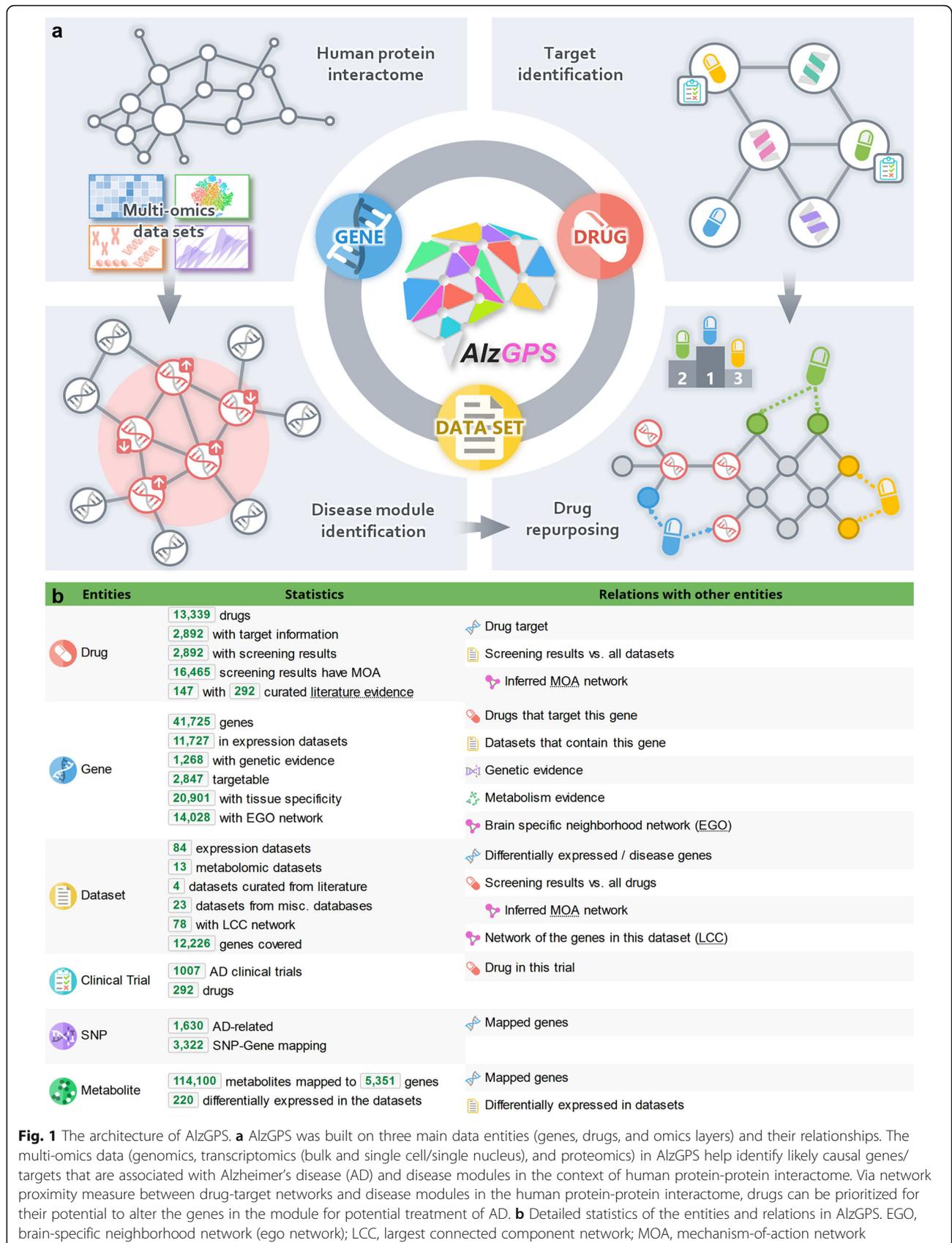
Background

Alzheimer's disease (AD) is a progressive neurodegenerative disorder accounting for 60–80% of dementia cases [1]. In addition to cognitive decline, AD patients have extensive neuropathological changes including deposition of extracellular amyloid plaques, intracellular neurofibrillary tangles, and neuronal death [2, 3]. It is estimated that the number of AD patients will reach 16 million by 2050 in the USA alone [4, 5]. Effective treatments are needed, as there are no disease-modifying treatments for AD and no new drugs have been approved since 2003 by the US Food and Drug Administration (FDA). There are several possible explanations for the high failure rate in AD drug discovery. For example, traditional “one gene, one drug, one disease” hypothesis may result in failure by anticipated off-target side effects and suboptimal efficacy because of complex disease pathobiology of AD [3, 6]. Also, there is a lack of sensitive endpoint measures for outcomes in clinical trials. Other potential immediate causes for clinical trial failures include targeting the wrong pathobiological or pathophysiological mechanisms, attempted intervention at the wrong stage (too early or too late), unfavorable pharmacodynamic and pharmacokinetic characteristics of the drug (e.g., poor brain penetration), lack of target engagement by drug candidates, and hypotheses that fail to incorporate the great complexity of AD [6, 7].

Multiple types of omics data have greatly facilitated our understanding of the pathobiology of AD. For example, using single-cell RNA-Seq, a novel microglia type (termed disease-associated microglia, DAM) was discovered to be associated with AD, understanding of whose molecular mechanism could offer new therapeutic targets [8]. Using large-scale genome-wide association studies (GWAS), twenty loci showed genome-wide significant association with AD, among which 11 were newly discovered [9]. A recent study using deep profiling of proteome and phosphoproteome prioritized proteins and pathways associated with AD, and it was shown that protein changes and their corresponding RNA levels only partially coincide [10]. The large amount of multi-omics data and recent advances in network-based methodologies for drug repurposing today present

unprecedented opportunities for accelerating target identification for drug discovery for AD. This potential has been demonstrated in other complex diseases as well, such as cancer [11], cardiovascular disease [12], and schizophrenia [13], and is beginning to be exploited in AD [6, 14]. Drug repurposing offers a rapid and cost-effective solution for drug discovery for complex disease, such as the current global pandemic of coronavirus disease 2019 (COVID-19) [15, 16] and AD [6]. The central idea of network-based drug repurposing is that for a drug to be able to affect a disease, the drug targets must directly overlap with or be in the immediate vicinity of the disease modules, which can be identified using the vast amount of high-throughput multi-omics data (Fig. 1a). Our recent efforts using network-based methodologies and AD omics data have led to the discovery of two drugs that show efficacy in network models in AD: sildenafil [6] and pioglitazone [14]. Network analysis provides potential mechanisms for these drugs and facilitates experimental validation. Therefore, we posit that a comprehensive systems biology tool in the framework of network-based multi-omics analysis could inform Alzheimer's patient care and therapeutic development.

To this end, we present a new freely available database and tool, named AlzGPS (A Genome-wide Positioning Systems platform for Alzheimer's Drug Discovery), for target identification and drug repurposing for AD. AlzGPS was built with large-scale diverse information, including multi-omics (genomics, transcriptomics [bulk and single cell], proteomics, and interactomics) of human and other species, drug-target networks, literature-derived evidence, AD clinical trials information, and network proximity analysis (Fig. 1b). Our hope is that AlzGPS will be a valuable resource for the AD research community for several reasons. First, AlzGPS contains abundant multi-domain information all coalesced in one location. The manually curated data, such as the literature-derived information for the most promising repurposable drugs and more than 100 multi-omics AD data sets, are of high quality and relevance. Second, using state-of-the-art network proximity approaches, AlzGPS provides a systemic evaluation of 3000 FDA approved or investigational drugs against the AD data sets.



These results (along with various network visualizations) will provide insights for potential repurposable drugs with clear network-based footprints in the context of the human protein-protein interactome. The drug-data set associations can be further explored in AlzGPS for individual drug targets or genes associated with AD. Lastly, AlzGPS offers a highly interactive and intuitive modern web interface. The relational nature of these data was embedded in the design to help the user easily navigate through different types of information. In addition, AlzGPS provides three types of network visualizations for the tens of thousands of networks in the database, including brain-specific neighbor networks for genes, disease modules derived from multi-omic profiles with varying degrees of disease biology of AD, and inferred mechanism-of-action (MOA) networks for drugs and omic pairs with significant proximity. AlzGPS is freely available to the public without registration requirement at <https://alzgps.lerner.ccf.org>.

Methods

Data collection and preprocessing

AD data sets

A data set is defined as either (1) genes/proteins/metabolites that are differentially expressed in AD patients/mice vs. controls, or (2) genes that have known associations with risks of AD from literature or other databases. We retrieved expression data sets underlying AD pathogenesis capturing transcriptomics (microarray, bulk or single-cell RNA-Seq) and proteomics across human, mouse, and model organisms (e.g., fruit fly and *Caenorhabditis elegans*). All the samples of the data sets were derived from total brain, specific brain regions (including hippocampus, cortex, and cerebellum), and brain-derived single cells, such as microglial cells. For some of the expression data sets, the differentially expressed genes/proteins were obtained from the original publications (from main tables or supplemental tables). For other data sets that did not have such differential expression results available, the original brain microarray/RNA-Seq data were obtained from Gene Expression Omnibus (GEO) [17] and differential expression analysis was performed using the tool GEO2R [18]. GEO2R performs the differential expression analysis for the sample groups defined by the user using the limma R package [19]. All differentially expressed genes identified in mouse were further mapped to unique human-orthologous genes using the NCBI HomoloGene database (<https://www.ncbi.nlm.nih.gov/homologene>). The details for all the data sets, including organism, genetic model (for mouse), brain region, cell type (for single-cell RNA-Seq), PubMed ID, GEO ID, and the sources (e.g., supplemental table or GEO2R), can be found in Table S1.

Genes and proteins

We retrieved the gene information from the HUGO Gene Nomenclature Committee (HGNC, <https://www.genenames.org/>) [20], including gene symbol, name, type (e.g., coding and non-coding), chromosome, synonyms, and identification (ID) mapping in various other databases such as NCBI Gene, ENSEMBL, and UniProt. All proteins from the AD proteomics data sets were mapped to genes using the mapping information from HGNC.

Single-nucleotide polymorphisms (SNPs)

We found 3321 AD-associated genetic records for 1268 genes mapped to 1629 SNPs, by combining results from GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) [21] using the trait “Alzheimer’s disease” and published studies. The PubMed IDs for the genetic evidence are provided in AlzGPS.

Tissue expression specificity

We downloaded RNA-Seq data (transcripts per million, TPM) across 33 human tissues from the GTEx v8 release (accessed on March 31, 2020, <https://gtexportal.org/home/>). We defined the genes with count per million (CPM) ≥ 0.5 in over 90% samples (e.g., brain) as tissue-expressed genes and otherwise as tissue-unexpressed. To quantify the expression significance of tissue-expressed gene i in tissue t , we calculated the average expression $\langle E(i) \rangle$ and the standard deviation $\delta_E(i)$ of a gene’s expression across all included tissues. The significance of gene expression in tissue t is defined as:

$$z_E(i, t) = \frac{E(i, t) - \langle E(i) \rangle}{\delta_E(i)} \quad (1)$$

Data for multiple brain regions were available from GTEx v8. We combined the data of these brain regions when comparing the brain expression specificity vs. other tissues. In addition, we further computed the expression specificity across 13 different brain regions. Both tissue expression specificity and brain region expression specificity results for the genes are available in AlzGPS.

Drugs

We retrieved drug information from the DrugBank database (v4.3) [22], including name, type, group (approved, investigational, etc.), Simplified Molecular-Input Line Entry System (SMILES), and Anatomical Therapeutic Chemical (ATC) code(s). We also evaluated the pharmacokinetic properties (such as blood–brain barrier [BBB]) of the drugs using admetSAR [23, 24].

Drug literature information for AD treatment

For the top 300 repurposable drugs (i.e., drugs with the highest number of significant proximities to the AD data sets), we manually searched and curated the literature for their therapeutic efficacy against AD using PubMed. In addition to the title, journal, and PubMed ID, we summarized the types (clinical and non-clinical), experimental settings (e.g., mouse/human and transgenic line for non-clinical studies; patient groups, randomization type, length, and control type of clinical studies), and results of these studies. In total, we found 292 studies for 147 drugs.

Drug-target network

To build a high-quality drug-target network, several databases were accessed, including the DrugBank database (v4.3) [22], Therapeutic Target Database (TTD) [25], PharmGKB database, ChEMBL (v20) [26], BindingDB [27], and IUPHAR/BPS Guide to PHAR MACOLOGY [28]. Only biophysical drug-target interactions involving human proteins were included. To ensure data quality, we kept only interactions that have inhibition constant/potency (K_i), dissociation constant (K_d), median effective concentration (EC_{50}), or median inhibitory concentration (IC_{50}) $\leq 10 \mu\text{M}$. The final drug-target network contains 21,965 interactions among 2892 drugs and 2847 human targets/proteins.

Clinical trials

The AD interventional clinical trials were retrieved from <https://clinicaltrials.gov>. Information including phase, posted date, status, and agent(s) was obtained from <https://clinicaltrials.gov>. Drugs were mapped to the DrugBank IDs. Proposed mechanism and therapeutic purpose were from Cummings et al. [29, 30].

Human protein interactome

We used our previously built high-quality comprehensive human protein interactome which contains 351,444 unique protein-protein interactions (PPIs, edges) among 17,706 proteins (nodes) [11, 12, 31, 32]. Briefly, five types of evidence were considered for building the interactome: physical PPIs from protein three-dimensional (3D) structures, binary PPIs revealed by high-throughput yeast-two-hybrid (Y2H) systems, kinase-substrate interactions by literature-derived low-throughput or high-throughput experiments, signaling networks by literature-derived low-throughput experiments, and literature-curated PPIs identified by affinity purification followed by mass spectrometry (AP-MS), Y2H, or by literature-derived

low-throughput experiments. The details are provided in our previous studies [11, 12, 31, 32].

Network proximity quantification of drugs and AD data sets

To quantify the associations between drugs and AD-related gene sets from the data sets, we adopted the “closest” network proximity measure:

$$\langle d_{AB} \rangle = \frac{1}{\|A\| + \|B\|} \left(\sum_{a \in A} \min_{b \in B} d(a, b) + \sum_{b \in B} \min_{a \in A} d(a, b) \right) \quad (2)$$

where $d(a, b)$ is the shortest path length between gene a and b from gene list A (drug targets) and B (AD genes), respectively. To evaluate whether such proximity was significant, we performed z score normalization using a permutation test of 1000 random experiments. In each random experiment, two randomly generated gene lists that have similar degree distributions to A and B were measured for the proximity. The z score was calculated as:

$$z_d = \frac{d - \bar{d}}{\sigma_d} \quad (3)$$

P value was calculated according to the permutation test. Drug-data set pairs with $Z < -1.5$ and $P < 0.05$ were considered significantly proximal. In addition to network proximity, we calculated two additional metrics, overlap coefficient C and Jaccard index J , to quantify the overlap and similarity of A and B :

$$C = \frac{|A \cap B|}{\min(|A|, |B|)} \quad (4)$$

$$J = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

Generation of gene/protein networks

We offer three types of networks in AlzGPS: brain-specific neighborhood (EGO) network for the genes, largest connected component (LCC) network for the data sets, and inferred MOA network for significantly proximal drug-data set pairs. The three networks differ by inclusion criteria of the nodes (genes/proteins). The edges are PPIs colored by their types (e.g., 3D, Y2H, and literature). All networks are colored by whether they can be targeted by the drugs in our database.

For the EGO networks, we filtered genes by their brain expression and generated only the network for those that were considered to be expressed in brain using GTEx data. We used the *ego_graph* function from NetworkX [33] to generate the EGO networks. The networks are centered around the genes-of-interest. We incorporated the tissue specificity of the genes (indicated

in the network by the node size) into the visualization tool, to allow users to further filter the network to show only the genes that have positive brain specificity.

An LCC network was generated for each AD data set using the *subgraph* function from networkx. For MOA, we examined the connections (PPIs) among the drug targets and the data sets.

Website implementation

AlzGPS was implemented with the Django v3.1.0 framework (www.djangoproject.com). The website frontend was implemented with HTML, CSS, and JavaScript. The frontend was designed to be highly interactive and integrative. It uses AJAX to asynchronously acquire data in JSON format based on user requests to dynamically update the frontend interface. This architecture can therefore be integrated into end users' own pipelines. Network visualizations were implemented using Cytoscape.js [34].

Results

Information architecture and statistics

One key feature of AlzGPS is the highly diverse yet interconnected data types (Fig. 1). The three main data types are genes, drugs, and AD-relevant omics data sets. More than 100 omics data sets were processed, including 84 expression data sets (Table S1) from AD transgenic animal models or patient-derived samples, 27 data sets from the literature or from other databases, and 13 metabolomic data sets. The expression data sets contain transcriptomic and proteomic data of human and rodent samples. Comparative sample groups were available in these data sets, such as early stage vs. late stage, healthy vs. AD. The differentially expressed genes/proteins were calculated for each data set.

The statistics and relations of the database are shown in Fig. 1b. We collected and processed all the basic information (see the "Methods" section) and then constructed the relationships among the data types. For example, for genes and drugs, the relationship is drugs targeting proteins (genes); for genes and data sets, the relationship is genes being differentially expressed in the expression data sets or included in other types of data sets, such as literature-based; for drugs and data sets, the proximity between each pair was calculated (see the "Methods" section) to identify the drug that is significantly proximal to a data set, and vice versa.

Additional data types were collected or generated. For genes, these included genetic evidence (variants associated with AD) and tissue expression specificity to provide additional information for target gene identification. For drugs, we collected the data from

more than 1000 AD clinical trials, and included the proposed mechanisms-of-action and possible therapeutic indications on AD [29, 30]. Drugs of these trials were extracted such that users can open associated drugs from the trial page. The BBB probability was computed [23, 24], as well as 23 other predicted absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties. For the top 300 drugs with the highest number of significant proximities to all the data sets, we manually curated the available literature. A total of 292 studies were found for 147 drugs (49%) that reported the associations of the drugs and AD. We grouped these studies into clinical and non-clinical, and extracted trial information for clinical type and experimental setting (number and type of patients) for both types. We also summarized and provided the study results.

Web interface and network visualizations

A highly interactive web interface was implemented (Fig. 2). On the home page (Fig. 2a), the user can search for drugs, genes, metabolites, gene variants, and clinical trials. The user can directly list all drugs by their first-level ATC code, all AD data sets available, and all the AD clinical trials (Fig. 2b). The search results are displayed in the "DATA TABLE" tab and switched with their associated buttons in the "RESULT" section on the left. Each data entity has its own data table for the associated information in the "DATA TABLE" tab. For example, on the gene page of APP (Fig. 2b) is the basic information (green rows), such as name, type, chromosome, and synonym; descriptions for the derived data (purple rows), such as tissue specificity and number of genetic records; and external links (red row). Data for the relations of APP and other entities can be loaded by clicking the button in "DETAIL" (blue row). For example, the expression data sets in which APP is differentially expressed can be found by clicking the "DATA-SET" button (Fig. 2b). Any data loaded will be added to the same explorer. The buttons in the "RESULT" are organized in trees. For example, APP is included in the "V1 AD-seed" data set, which contains 144 AD-associated genes with strong literature evidence. When the user clicks this data set in the APP gene table, a new data table for the "V1 AD-seed" data set will replace the APP gene page, and a new button with indentation will appear below the APP button in "RESULT" (Fig. 2b).

The all-in-one interactive explorer that minimizes the need for navigation of information using the relational nature of these data is a major feature of the web interface. Another major feature is the network visualizations. We offer three types of networks: (1) the brain-specific neighborhood network (EGO) for a gene-of-interest that shows the PPIs with its neighbors (Fig. 2c), (2) the

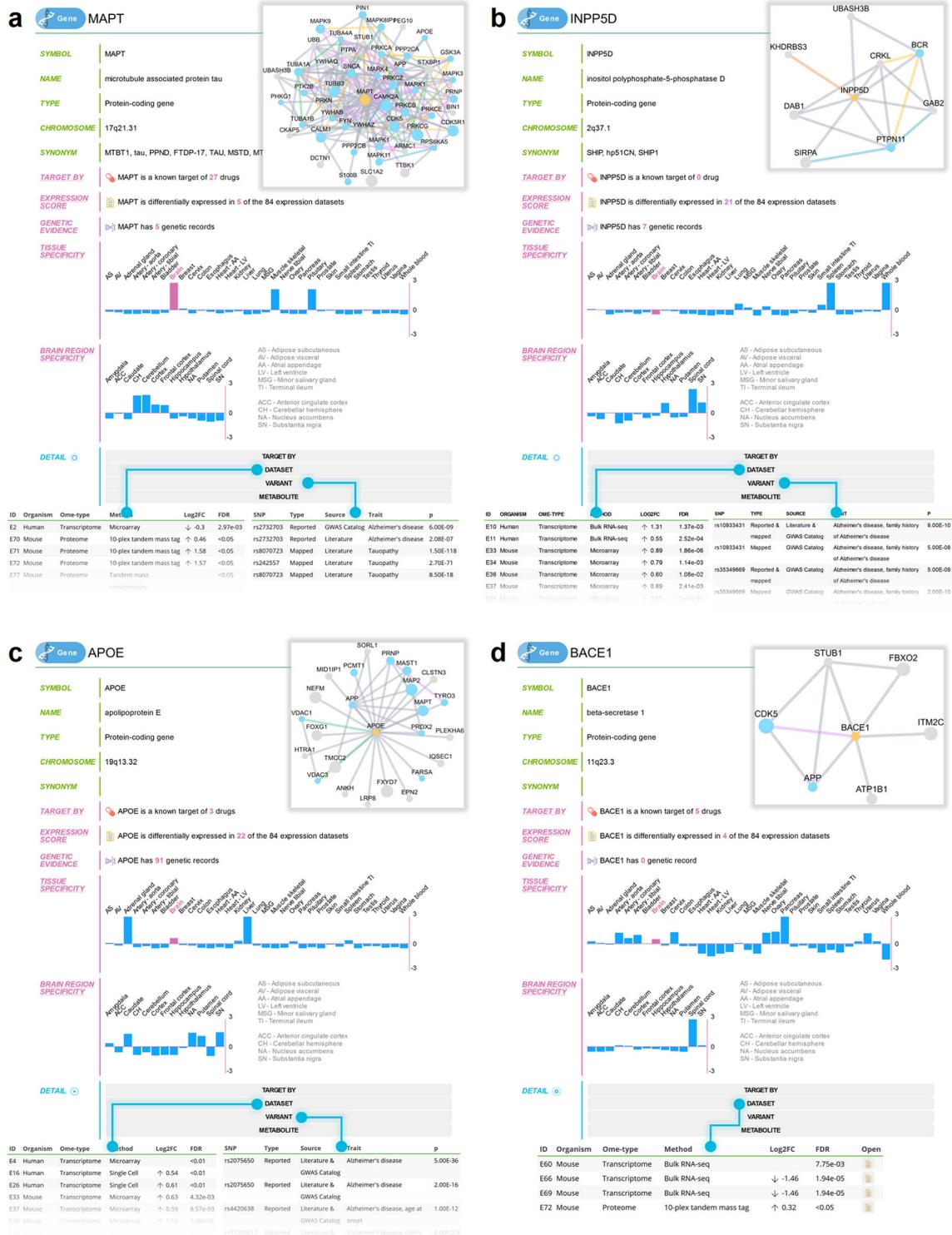


Fig. 3 Case study—target identification. Four genes, MAPT (a), INPP5D (b), APOE (c), and BACE1 (d), are used as examples to show the gene page. On the gene page, we show a summary of several statistics of the gene in AlzGPS, including the number of drugs that can target it, number of data sets of omics in which the target/protein coding gene is differentially expressed, number of genetic records, and the brain expression specificity. Detailed information can be loaded by clicking corresponding buttons. Examples of detailed differential expression results and genetic records are shown for these four genes. In addition, a brain-specific neighborhood network is available that centers around the gene-of-interest and shows the targetability of its neighborhood

largest connected component (LCC) network for a data set that shows the largest module formed by the genes in this data set (Fig. 2d), and (3) inferred MOA network for a significantly proximal drug-data set pair, which is illustrated in the case studies below.

Case study—target identification

Generally, using AlzGPS for AD target identification starts with selecting one or a set of data sets (Fig. 2b, “DATASET” tab). Users can select a data set based on organisms, methods (e.g., single-cell/nuclei RNA-Seq), brain regions, and comparisons (e.g., early-onset AD vs. healthy control) for the expression data sets. Additionally, we have collected data sets from the literature, other databases, or computationally predicted results. Here, we use the “V1 AD-seed” data set as a starting point. This data set was from our recent study which contains 144 AD-associated genes based on literature-derived evidence. We found that 118 genes were differentially expressed as shown in at least one data set. By browsing these genes, we selected four examples, microtubule-associated protein tau (MAPT), inositol polyphosphate-5-phosphatase D (INPP5D), apolipoprotein E (APOE), and β -secretase 1 (BACE1) based on positive brain expression specificity and number of data sets that include them.

MAPT

MAPT encodes the tau protein, modification of which is one of the main neuropathological hallmarks of AD [35, 36]. Mutations and alternative splicing of MAPT are associated with risk of AD [37]. MAPT is differentially expressed in five expression data sets (Fig. 3a) and has high brain specificity. Five pieces of genetic evidence were found for MAPT. MAPT can be targeted by 27 drugs. In addition, many of its direct PPI neighbors are targetable, suggesting a potential treatment strategy by targeting MAPT and its neighbors.

INPP5D

We found 7 genetic association records for INPP5D (Fig. 3b). Recent GWAS results showed that the rs35349669 polymorphism of INPP5D was significantly associated with an increased risk of late-onset AD in Caucasians [9, 38]. The intronic SNP rs61068452 of INPP5D was significantly associated with reduced cerebrospinal fluid (CSF) t-tau/A β _{1–42} ratio, showing a potentially protective role in AD [39]. In addition to these genetic associations, INPP5D was also differentially manifested across 21 human and mouse expression data sets. Altogether, INPP5D may suggest potential drug target candidates for future therapeutic development.

APOE

APOE has three major alleles, ϵ 2, ϵ 3, and ϵ 4. Individuals carrying the ϵ 4 allele have an increased risk of developing AD compared to those carrying the more common ϵ 3 allele, while ϵ 2 decreases the risk [40, 41]. The ϵ 4 allele of APOE is the main genetic risk factor of AD [41]. APOE ϵ 4 plays an important role in A β deposition [41], a major pathological hallmark of AD. APOE is differentially expressed in 22 data sets (Fig. 3c). It has a high number of associated genetic records—91. Both APOE and its PPI partners can be targeted.

BACE1

BACE1 cleaves APP and generates A β peptides [42], whose aggregation is a pathological hallmark of AD. The inhibition of BACE1 has been a popular target for AD drug development. Shown in Fig. 3d, BACE1 is differentially expressed in 4 data sets.

Case study—drug repurposing

In this section, we use sildenafil and pioglitazone as two examples. In our recent studies, we found that both sildenafil and pioglitazone were associated with a reduced risk of AD using network proximity analysis and retrospective case-control validation [14]. Mechanistically, in vitro assays showed that both drugs were able to downregulate cyclin-dependent kinase 5 (CDK5) and glycogen synthase kinase 3 beta (GSK3B) in human microglia cells. These drugs were discovered using different data sets. Sildenafil was found using a high-quality literature-based AD endophenotype module (available as AlzGPS data set “V1 AD-seed”) containing 144 genes. Pioglitazone was found using 103 high-confidence AD risk genes (available as AlzGPS data set “V4 AD-inferred-GWAS-risk-genes”) identified by GWAS [13].

AlzGPS provides a list-view of the network proximity results of all the drugs organized by their first-level ATC code, which can be found in the “DRUG CLASS” tab (Fig. 2b). The drugs are ranked by the number of significant proximities to the data sets. Sildenafil is in the top four of the 148 drugs under the ATC code G “Genitourinary system and sex hormones” with network proximity results, the top three being vardenafil, ibuprofen, and gentian violet cation. Pioglitazone is in the top six of the 226 drugs under the ATC code A “Alimentary tract and metabolism,” following tetracycline, human insulin, epinephrine, cholecalciferol, and teduglutide. Both drugs achieved high numbers of significant proximities to the expression data set. Next, we examined the basic information of these drugs (Fig. 4a, e). Both drugs are predicted to be BBB penetrable. Sildenafil has 20 known targets and is significantly proximal to 27 of the 111 data sets (Fig. 4a). We found one non-clinical study that reported that sildenafil treatment improves cognition and

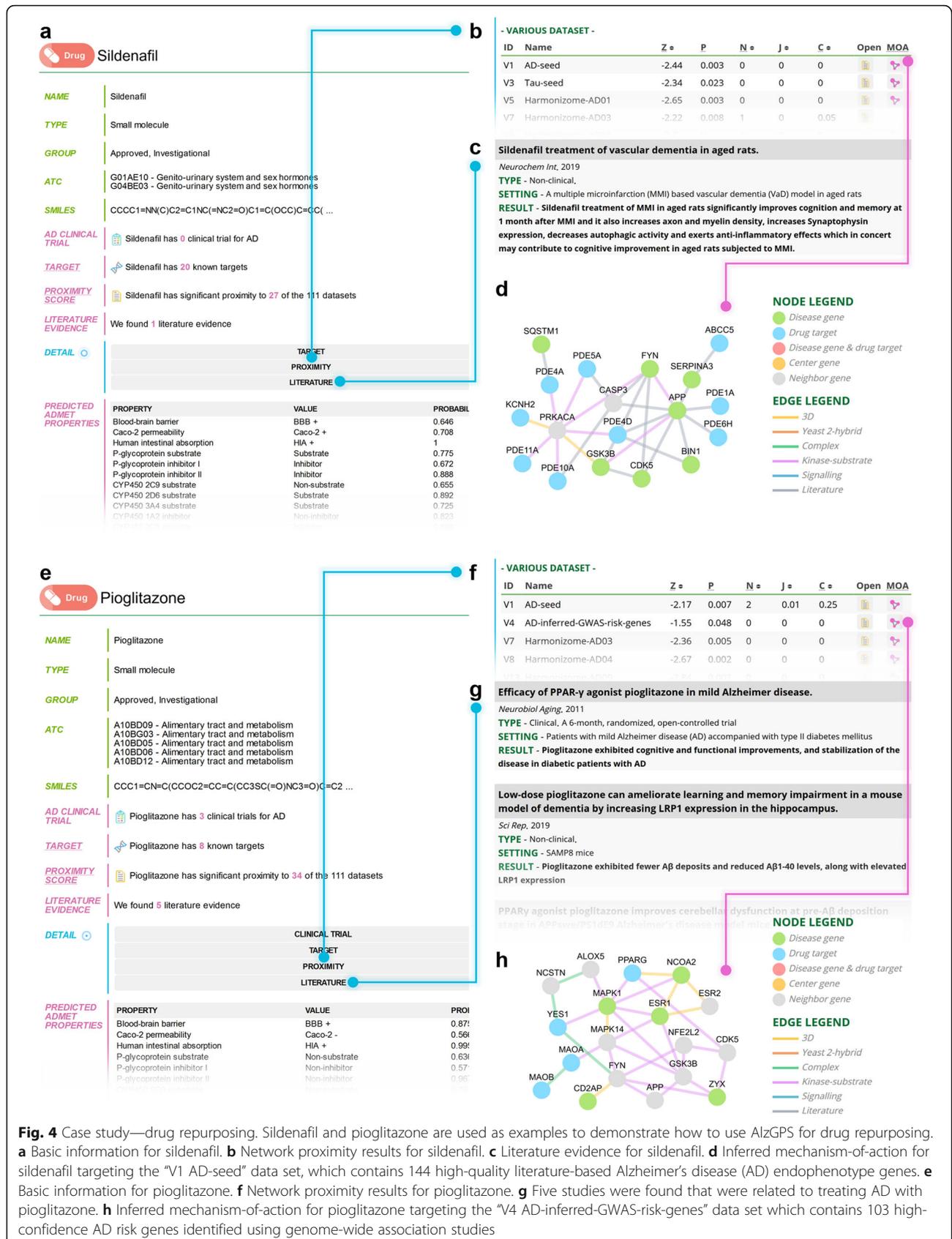


Fig. 4 Case study—drug repurposing. Sildenafil and pioglitazone are used as examples to demonstrate how to use AlzGPS for drug repurposing. **a** Basic information for sildenafil. **b** Network proximity results for sildenafil. **c** Literature evidence for sildenafil. **d** Inferred mechanism-of-action for sildenafil targeting the “V1 AD-seed” data set, which contains 144 high-quality literature-based Alzheimer’s disease (AD) endophenotype genes. **e** Basic information for pioglitazone. **f** Network proximity results for pioglitazone. **g** Five studies were found that were related to treating AD with pioglitazone. **h** Inferred mechanism-of-action for pioglitazone targeting the “V4 AD-inferred-GWAS-risk-genes” data set which contains 103 high-confidence AD risk genes identified using genome-wide association studies

memory of vascular dementia in aged rats [43] (Fig. 4c). As noted, we identified the potential of sildenafil against AD using the AD endophenotype module (Fig. 4b, $Z = -2.44$, $P = 0.003$). Then, clicking the corresponding “MOA (mechanism-of-action)” button opened the inferred MOA network for sildenafil and the data set (Fig. 4d). Although sildenafil does not target the genes in the data set (green) directly, it can potentially alter them through PPIs with its targets (blue).

Pioglitazone has 8 known targets and is significantly proximal to 34 data sets (Fig. 4e). Five studies containing both clinical and non-clinical data were found to be related to treating AD with pioglitazone. For example, a clinical study showed that pioglitazone can improve cognition in AD patients with type II diabetes [44] (Fig. 4g). Similarly, network results and associated MOA networks suggested that pioglitazone can affect AD risk genes through PPIs (Fig. 4f, h).

Validation studies

Once candidate agents are identified on AlzGPS, a variety of validation steps can be pursued [6]. The agent can be tested in animal model systems of AD pathology to evaluate the predicted MOA of behavioral and biological effects. Since these are repurposed agents and have been used for other indications in human healthcare, electronic medical records can be interrogated to determine if there are notable effects on AD incidence, prevalence, or rate of progression. Both these methods are imperfect since animal models have rarely been predictive of human response, and doses and duration of exposures may be different for indications other than AD in which the candidate agents are used. The ultimate assessment that could make an agent available for human care is success in a clinical trial and nominated agents must eventually be submitted to trials. If repurposed agents are not entered into trials because of intellectual property limitations or other challenges, the information from AlzGPS may be useful in identifying druggable disease pathways or providing seed structures that provide a basis for creation of related novel agents with similar MOAs.

Discussion

Dr. Alois Alzheimer first described the condition in 1907, but scientists have not been able to develop any disease-modifying treatments for AD since. In this study, we developed a computational platform, termed AlzGPS (<https://alzgps.lerner.ccf.org>), which will advance genome-informed Alzheimer's patient care and therapeutic development, by leveraging all existing multi-omics knowledge and data. To be specific, AlzGPS enables searching, sharing, visualizing, querying, and analyzing multi-omics (genomics, transcriptomics,

proteomics, metabolomics, and interactomics), different types of heterogeneous bio-networks, and clinical databases for genome-informed target identification and drug repurposing for potential treatment of AD. In addition, drug candidates prioritized by AlzGPS may offer possible tool compounds for investigation of disease biology or pathobiology of AD. We believe that AlzGPS will be a valuable tool for the AD drug discovery community by providing (1) (manually curated) abundant diverse information of AD multi-omics data sets, genes, and drugs; (2) drug repurposing results using state-of-the-art network proximity approaches for novel insights; and (3) highly interactive and intuitive web interface with informative network visualizations. To the best knowledge of the authors, this study presents the first AD multi-omics framework using both network-based methodologies and genome-informed precision medicine drug discovery for AD.

We acknowledge several potential limitations in current AlzGPS. First, we assembled multi-omics data and clinical trial data from diverse sources. Data harmonization is a crucial issue which should be addressed in the future, possibly through machine learning approaches. Second, although we assembled comprehensive PPIs based on our sizeable efforts, incompleteness of human protein-protein interactome data and potential literature bias may influence performance of AlzGPS. For example, well-studied genes (such as *APOE*, *MAPT*, and *BACE1*) are top prioritized target candidates as they have more accumulating PPIs, genetic and genomic information. In addition, the current implementation of AlzGPS does not differentiate the allele-specific expression. We will integrate more isoform-specific expression profiles (such as *APOE* $\epsilon 2/\epsilon 3/\epsilon 4$) in the AlzGPS in the future. Although we integrated large-scale genetic data from meta-analyses of GWAS, whole-genome/exome sequencing data for AD are missing in current AlzGPS. We will integrate high-throughput next-generation DNA sequencing from multiple national AD genome projects, including the Alzheimer's Disease Sequencing Project (ADSP) and the Alzheimer's Disease Neuroimaging Initiative (ADNI), especially DNA/RNA sequencing from minority population, which will provide unbiased genomic resources to prioritize novel drug targets. Systematic evaluation of pharmacokinetic properties (including brain penetration) for drugs using in silico approaches and publicly available in vitro and in vivo assays is highly encouraged in the future. Finally, we will integrate clinical trial and approved drug information from other sources, including European Medicines Agency, Pharmaceuticals and Medical Devices Agency at Japan, and the China Food and Drug Administration, to advance the international

Alzheimer's research communities under the AlzGPS framework. We will continue to add more types of omics data and update AlzGPS annually or when a large amount of new data is available.

Conclusions

In summary, AlzGPS presents the first comprehensive *in silico* tool for human genome-informed precision medicine drug discovery for AD. AlzGPS contains rich and diverse information connecting genetics, genomics, proteomics, and metabolomics for disease pathobiology, and drugs for AD target identification and drug repurposing. It utilizes multiple biological networks and omics data, and provides network-based drug repurposing results with network visualizations. From a translational perspective, if broadly applied, AlzGPS will offer a powerful tool for prioritizing biologically relevant targets and clinically relevant repurposed drug candidates and tool compounds for multi-omics-informed discovery in AD and other neurodegenerative diseases.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13195-020-00760-w>.

Additional file 1: Table S1. All data sets in AlzGPS.

Acknowledgements

We thank the Lerner Research Institute Computing Services for hosting AlzGPS.

Authors' contributions

F.C. conceived the study. Y.Z. constructed the database and developed the website. J.F., Y.Z., and Y.H.K. performed the data gathering and processing. L.M.B., A.A.P., J.B.L., and J.C. discussed and interpreted all results. Y.Z., F.C., and J.C. wrote and all authors critically revised the manuscript and gave final approval.

Funding

This work was supported by the National Institute of Aging (NIA) under Award Number R01AG066707 and 3R01AG066707-01S1 to F.C. This work was supported in part by the NIA under Award Number R56AG063870 to F.C., L.M.B., and J.B.L., A.A.P., L.M.B., J.C., J.B.L., and F.C. are supported together by the Translational Therapeutics Core of the Cleveland Alzheimer's Disease Research Center (NIH/NIA: 1 P30 AG062428-01). A.A.P. is also supported by the Brockman Foundation, Project 19PABH134580006-AHA/Allen Initiative in Brain Health and Cognitive Impairment, the Elizabeth Ring Mather & William Gwinn Mather Fund, S. Livingston Samuel Mather Trust, G.R. Lincoln Family Foundation, Wick Foundation, Gordon & Evie Safran, the Leonard Krieger Fund of the Cleveland Foundation, the Maxine and Lester Stoller Parkinson's Research Fund, and Louis Stokes VA Medical Center resources and facilities. Dr. Leverenz is supported by the Alzheimer's Drug Discovery Foundation, Cleveland Clinic Lerner Research Institute, Department of Defense, Douglas Herthel DVM Memorial Research Fund, Eisai, GE Healthcare, Jane and Lee Seidman Fund, Lewy Body Dementia Association, Michael J Fox Foundation, NIH/NIA funds (P30 AG062428, UO1 NS100610, R01 AG022304, R01 AG0577552, R03 AG063235, R21 AG064271, P20 AG068053), and Sanofi. Dr. Cummings is supported by Keep Memory Alive (KMA); NIGMS grant P20GM1109025; NINDS grant U01NS093334; and NIA grant R01AG053798.

Availability of data and materials

All the data in AlzGPS can be freely accessed without registration requirement at <https://alzgps.lerner.ccf.org>.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

Dr. Cummings has provided consultation to Acadia, Actinogen, Alkahest, Alzheon, Annovis, Avanir, Axsome, Biogen, BioXcel, Cassava, Cerecin, Cerevel, Cortexyme, Cytox, EIP Pharma, Eisai, Foresight, GemVax, Genentech, Green Valley, Grifols, Karuna, Merck, Novo Nordisk, Otsuka, Resverlogix, Roche, Samumed, Samus, Signant Health, Suven, Third Rock, and United Neuroscience pharmaceutical and assessment companies. Dr. Cummings has stock options in ADAMAS, AnnovisBio, MedAvante, and BiOasis. Dr. Leverenz has received consulting fees from Acadia, Biogen, Eisai, GE Healthcare, and Sunovion. The other authors have declared no competing interest.

Author details

¹Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA. ²Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA. ³Harrington Discovery Institute, University Hospitals Cleveland Medical Center, Cleveland, OH 44106, USA. ⁴Department of Psychiatry, Case Western Reserve University, Cleveland, OH 44106, USA. ⁵Geriatric Psychiatry, GRECC, Louis Stokes Cleveland VA Medical Center, Cleveland, OH 44106, USA. ⁶Institute for Transformative Molecular Medicine, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA. ⁷Weill Cornell Autism Research Program, Weill Cornell Medicine of Cornell University, New York, NY 10065, USA. ⁸Department of Neuroscience, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA. ⁹Lou Ruvo Center for Brain Health, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, USA. ¹⁰Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV 89106, USA. ¹¹Chambers-Grundy Center for Transformative Neuroscience, Department of Brain Health, School of Integrated Health Sciences, UNLV, Las Vegas, NV 89154, USA. ¹²Case Comprehensive Cancer Center, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA.

Received: 17 September 2020 Accepted: 23 December 2020

Published online: 13 January 2021

References

- 2020 Alzheimer's disease facts and figures. *Alzheimers Dement.* 2020;16(3):391–460.
- Long JM, Holtzman DM. Alzheimer disease: an update on pathobiology and treatment strategies. *Cell.* 2019;179(2):312–39.
- Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. Alzheimer's disease. *Nat Rev Dis Primers.* 2015;1:15056.
- Kodamullil AT, Zekri F, Sood M, Hengerer B, Canard L, McHale D, et al. Trial watch: tracing investment in drug development for Alzheimer disease. *Nat Rev Drug Discov.* 2017;16(12):819.
- Alteri E, Guizzaro L. Be open about drug failures to speed up research. *Nature.* 2018;563(7731):317–9.
- Fang J, Pieper AA, Nussinov R, Lee G, Bekris L, Leverenz JB, et al. Harnessing endophenotypes and network medicine for Alzheimer's drug repurposing. *Med Res Rev.* 2020;40(6):2386–426.
- Cummings J, Feldman HH, Scheltens P. The "rights" of precision drug development for Alzheimer's disease. *Alzheimers Res Ther.* 2019;11(1):76.
- Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell.* 2017;169(7):1276–90 e17.
- Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45(12):1452–8.
- Bai B, Wang X, Li Y, Chen PC, Yu K, Dey KK, et al. Deep multilayer brain proteomics identifies molecular networks in Alzheimer's disease progression. *Neuron.* 2020;105(6):975–91 e7.
- Cheng F, Lu W, Liu C, Fang J, Hou Y, Handy DE, et al. A genome-wide positioning systems network algorithm for *in silico* drug repurposing. *Nat Commun.* 2019;10(1):3476.

12. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabasi AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun.* 2018;9(1):2691.
13. Wang Q, Chen R, Cheng F, Wei Q, Ji Y, Yang H, et al. A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. *Nat Neurosci.* 2019;22(5):691–9.
14. Fang J, Zhang P, Wang Q, Zhou Y, Chiang WC, Cheng R, et al. Network-based translation of GWAS findings to pathobiology and drug repurposing for Alzheimer's disease. medRxiv. 2020:2020.01.15.20017160. <https://doi.org/10.1101/2020.01.15.20017160>.
15. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 2020;6:14.
16. Zhou Y, Wang F, Tang J, Nussinov R, Cheng F. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit Health.* 2020;2(12):e667–e76.
17. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30(1):207–10.
18. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013;41(Database issue):D991–5.
19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
20. Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res.* 2019;47(D1):D786–D92.
21. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005–D12.
22. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 2014;42(Database issue):D1091–7.
23. Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, et al. admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. *J Chem Inf Model.* 2012;52(11):3099–105.
24. Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, et al. Correction to “admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties”. *J Chem Inf Model.* 2019;59(11):4959.
25. Yang H, Qin C, Li YH, Tao L, Zhou J, Yu CY, et al. Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res.* 2016;44(D1):D1069–74.
26. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 2012;40(Database issue):D1100–7.
27. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 2007;35(Database issue):D198–201.
28. Pawson AJ, Sharman JL, Benson HE, Faccenda E, Alexander SP, Buneman OP, et al. The IUPHAR/BPS Guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. *Nucleic Acids Res.* 2014;42(Database issue):D1098–106.
29. Cummings J, Lee G, Ritter A, Zhong K. Alzheimer's disease drug development pipeline: 2018. *Alzheimers Dement (N Y).* 2018;4:195–214.
30. Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2019. *Alzheimers Dement (N Y).* 2019;5:272–93.
31. Cheng F, Kovacs IA, Barabasi AL. Network-based prediction of drug combinations. *Nat Commun.* 2019;10(1):1197.
32. Smith IN, Thacker S, Seyfi M, Cheng F, Eng C. Conformational dynamics and allosteric regulation landscapes of germline PTEN mutations associated with autism compared to those associated with cancer. *Am J Hum Genet.* 2019;104(5):861–78.
33. Hagberg A, Schult D, Swart P. Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy2008)*; 2008.
34. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics.* 2016;32(2):309–11.
35. Kosik KS, Joachim CL, Selkoe DJ. Microtubule-associated protein tau (tau) is a major antigenic component of paired helical filaments in Alzheimer disease. *Proc Natl Acad Sci U S A.* 1986;83(11):4044–8.
36. Goedert M, Wischik CM, Crowther RA, Walker JE, Klug A. Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau. *Proc Natl Acad Sci U S A.* 1988;85(11):4051–5.
37. Kovacs GG. Invited review: neuropathology of tauopathies: principles and practice. *Neuropathol Appl Neurobiol.* 2015;41(1):3–23.
38. Jing H, Zhu JX, Wang HF, Zhang W, Zheng ZJ, Kong LL, et al. INPP5D rs35349669 polymorphism with late-onset Alzheimer's disease: a replication study and meta-analysis. *Oncotarget.* 2016;7(43):69225–30.
39. Yao X, Risacher SL, Nho K, Saykin AJ, Wang Z, Shen L, et al. Targeted genetic analysis of cerebral blood flow imaging phenotypes implicates the INPP5D gene. *Neurobiol Aging.* 2019;81:213–21.
40. Mahley RW, Rall SC Jr. Apolipoprotein E: far more than a lipid transport protein. *Annu Rev Genomics Hum Genet.* 2000;1:507–37.
41. Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol.* 2013;9(2):106–18.
42. Cai H, Wang Y, McCarthy D, Wen H, Borchelt DR, Price DL, et al. BACE1 is the major beta-secretase for generation of Abeta peptides by neurons. *Nat Neurosci.* 2001;4(3):233–4.
43. Venkat P, Chopp M, Zacharek A, Cui C, Landschoot-Ward J, Qian Y, et al. Sildenafil treatment of vascular dementia in aged rats. *Neurochem Int.* 2019;127:103–12.
44. Sato T, Hanyu H, Hirao K, Kanetaka H, Sakurai H, Iwamoto T. Efficacy of PPAR-gamma agonist pioglitazone in mild Alzheimer disease. *Neurobiol Aging.* 2011;32(9):1626–33.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

