RESEARCH ARTICLE

The Alzheimer's Cell Atlas (TACA): A single-cell molecular map for translational therapeutics accelerator in Alzheimer's disease

Yadi Zhou[1] | **Jielin Xu**[1] | **Yuan Hou**[1] | **Lynn Bekris**[1,2] | **James B. Leverenz**[2,3] | **Andrew A. Pieper**[4,5,6,7,8] | **Jeffrey Cummings**[9] | **Feixiong Cheng**[1,2,10]

[1]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA

[2]Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

[3]Lou Ruvo Center for Brain Health, Neurological Institute, Cleveland Clinic, Cleveland, Ohio, USA

[4]Harrington Discovery Institute, University Hospitals Cleveland Medical Center, Cleveland, Ohio, USA

[5]Department of Psychiatry, Case Western Reserve University, Cleveland, Ohio, USA

[6]Geriatric Psychiatry, GRECC, Louis Stokes Cleveland VA Medical Center, Cleveland, Ohio, USA

[7]Institute for Transformative Molecular Medicine, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

[8]Department of Neuroscience, Case Western Reserve University, School of Medicine, Cleveland, Ohio, USA

[9]Chambers-Grundy Center for Transformative Neuroscience, Pam Quirk Brain Health and Biomarker Laboratory, Department of Brain Health, School of Integrated Health Sciences, University of Nevada Las Vegas, Las Vegas, Nevada, USA

[10]Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA

Correspondence
Feixiong Cheng, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA.
Email: chengf@ccf.org

Abstract

Introduction: Recent advances in generating massive single-cell/nucleus transcriptomic data have shown great potential for facilitating the identification of cell type–specific Alzheimer's disease (AD) pathobiology and drug-target discovery for therapeutic development.

Methods: We developed The Alzheimer's Cell Atlas (TACA) by compiling an AD brain cell atlas consisting of over 1.1 million cells/nuclei across 26 data sets, covering major brain regions (hippocampus, cerebellum, prefrontal cortex, and so on) and cell types (astrocyte, microglia, neuron, oligodendrocytes, and so on). We conducted nearly 1400 differential expression comparisons to identify cell type–specific molecular alterations (e.g., case vs healthy control, sex-specific, apolipoprotein E (*APOE*) ε4/ε4, and TREM2 mutations). Each comparison was followed by protein-protein interaction module detection, functional enrichment analysis, and omics-informed target and drug (over 700,000 perturbation profiles) screening. Over 400 cell-cell interaction analyses using 6000 ligand-receptor interactions were conducted to identify the cell-cell communication networks in AD.

Yadi Zhou, Jielin Xu, and Yuan Hou contributed equally to this work.

Results: All results are integrated into TACA (https://taca.lerner.ccf.org/), a new web portal with cell type–specific, abundant transcriptomic information, and 12 interactive visualization tools for AD.

Discussion: We envision that TACA will be a highly valuable resource for both basic and translational research in AD, as it provides abundant information for AD pathobiology and actionable systems biology tools for drug discovery.

Highlights

- We compiled an Alzheimer's disease (AD) brain cell atlas consisting of more than 1.1 million cells/nuclei transcriptomes from 26 data sets, covering major brain regions (cortex, hippocampus, cerebellum) and cell types (e.g., neuron, oligodendrocyte, astrocyte, and microglia).
- We conducted over 1400 differential expression (DE) comparisons to identify cell type–specific gene expression alterations. Major comparison types are (1) AD versus healthy control; (2) sex-specific DE, (3) genotype-driven DE (i.e., apolipoprotein E [*APOE*] $\varepsilon 4/\varepsilon 4$ vs *APOE* $\varepsilon 3/\varepsilon 3$; TREM2^{R47H} vs common variants) analysis; and (4) others. Each comparison was further followed by (1) human protein-protein interactome network module analysis, (2) pathway enrichment analysis, and (3) gene-set enrichment analysis.
- For drug screening, we conducted gene set enrichment analysis for all the comparisons with over 700,000 drug perturbation profiles connecting more than 10,000 human genes and 13,000 drugs/compounds.
- A total of over 400 analyses of cell-cell interactions against 6000 experimentally validated ligand-receptor interactions were conducted to reveal the disease-relevant cell-cell communications in AD.

1 | INTRODUCTION

Alzheimer's disease (AD) is a devastating neurodegenerative disease now affecting 6.5 million Americans age 65 and older and projected to double to 13.8 million by 2060.[1] More than 11 million family members and unpaid caregivers provided an estimated $271.6 billion care to people with AD and other dementias in 2021,[1] while the attrition rate for AD clinical trials (2002–2012) is estimated at over 99%.[2] The underlying disease etiology and molecular mechanisms of AD are under investigation.[3–6] The genetic predisposition to AD involves a complex, polygenic, and pleiotropic genetic architecture.[7,8] The traditional reductionist paradigm overlooks the inherent complexity of AD and often leads to incomplete evidence on disease initiation, progression, or modification.[9] Existing multi-omics data resources, including genomics, transcriptomics, proteomics, and interactomics (protein-protein interactions [PPIs]), have not been fully utilized and integrated to identify pathobiology and support therapeutic development for AD and AD-related dementias (ADRDs). For example,

tools such as Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) and CELLxGENE (https://cellxgene.cziscience.com/) have an extensive number of single-cell/nucleus (sc/sn) omic data sets. These tools focus on visualizing cells (annotations) and gene expressions, but have not utilized resources such as PPIs to reveal underlying disease pathobiology and actionable targets, or utilized drug perturbation profiles for therapeutic discoveries. It is urgent to develop genome-wide, systems approaches or resources to identify likely molecular drivers and disease networks, which will enable a more complete mechanistic understanding of AD/ADRDs and assist in identifying effective treatments.[10–12]

Recent breakthroughs in sc/sn RNA-sequencing (RNA-seq) technologies have advanced our understanding of AD/ADRDs.[13,14] For example, using 5XFAD mouse model scRNA-seq data, a novel microglia subtype termed disease-associated microglia (DAM) was discovered that co-localized with amyloid beta (Aβ) plaques.[13] Diseased-associated astrocyte (DAA) was also discovered using a snRNA-seq data set, which occurred in AD mouse models and increased with

disease progression.[14] Using a large-scale human snRNA-seq data set, researchers discovered two distinct microglial subclusters in patients with AD that correlated with Aβ plaques and tau pathology, respectively.[15] Marked disease heterogeneity of AD may have been one of the leading causes of the high failure rate of AD clinical trials.[16] These sc/sn studies advance our understanding of the heterogeneity of AD and offer cell type–specific actionable targets and, therefore, have great potential in target identification and precision-medicine drug repurposing for AD.[10,12,17] For example, using endophenotype network and population-based validation, we identified that sildenafil use was associated significantly with a 69% reduced likelihood of AD, potentially by promoting neurite growth and decreasing phospho-tau expression in patients with AD.[10] Using sc/sn RNA-seq data and network-based methodologies, our team identified both unique and shared immune pathways between DAM and astrocytes, and performed network-based predictions that identified fluticasone as a potential treatment for AD.[17]

Although there has been a surge of new AD-related sc/sn RNA-seq data sets in the past few years,[13–15,18–31] the potential insights embedded in these data come with several difficulties. The majority of the original studies of these heterogeneous data sets focus on specific aspects of AD, although some studies such as Mathys et al.[32] and Grubman et al.[19] provide a comprehensive view of the AD biology in cell type–specific manners. Researchers frequently need to re-run the single-cell analysis pipelines for their tasks due to limited access to processed data and results, and such analyses require a large amount of computing resources. The application of state-of-the-art techniques, such as network pathobiology mapping, have been lacking with these data sets. To overcome these limitations, we built The Alzheimer's Cell Atlas (TACA), which contains abundant AD-related sc/sn transcriptomic information and various types of large-scale transcriptomic and systems biology analysis results for the identification of cell type–specific AD pathobiology and target discovery for rapid translational therapeutic development (e.g., drug repurposing).

2 | METHODS

The construction of TACA involved three steps: data collection (Figure 1A), data analysis (Figure 1B), and construction and implementation of the database and web portal (Figures 1C and 2A). Detailed methods can be found in the Supplementary Methods. In this initial version of TACA, three interactive explorers were implemented for genes (Figure 2B), drugs (Figure 2C), and sc/sn data sets (Figures 1C and 2D), respectively. A total of 12 visualization tools were implemented, among which seven are for different types of network visualizations.

3 | RESULTS

3.1 | Overall design of TACA

In this study, we compiled an AD brain sc/sn atlas consisting of more than 1.1 million cells/nuclei from over 400 human/mouse samples

RESEARCH IN CONTEXT

1. **Systematic review**: We reviewed the literatures using traditional sources (i.e., PubMed) and we have seen a surge in the number of Alzheimer's disease (AD) single-cell/nucleus multi-omics data sets in the past few years. Yet, genome-wide, systems biology approaches or resources that utilize these large-scale data to identify likely molecular drivers, disease networks, and drug target are still lacking. The development of a portal for these analyses results will enable a more complete mechanistic understanding of Alzheimer's disease (or AD) and assist in identifying treatments.

2. **Interpretation**: We compiled an AD brain cell atlas (termed The Alzheimer's Cell Atlas [TACA], https://taca.lerner.ccf.org/) consisting of more than 1 million cells/nuclei from 26 data sets, covering major brain regions (cortex, hippocampus, cerebellum, and so on) and cell types (neuron, oligodendrocyte, astrocyte, microglia, and so on). We developed a web portal with 12 interactive visualization tools (including cells, targets, drugs, and networks) and databases incorporating large-scale single cell/nucleus transcriptomic, various biological networks, and analyses results to facilitate the identification of cell type–specific AD pathobiology and drug-target identification for therapeutic discovery.

3. **Future directions**: We envision that TACA will be a highly valuable resource for both basic and translational research for AD, owing to the abundant information it contains for the AD pathobiology and the actionable systems biology tools it is equipped with for therapeutic discovery. We will continue to bring more single cell/nucleus transcriptomic data and more types of analyses results and visualizations into TACA.

across 26 data sets (Tables S1 and S2). All data sets were processed in consistent pipelines. We exhaustively compared gene expression among groups by automating the differential expression (DE) analyses using metadata that we curated from the original studies, reaching 1400 comparisons (Table S3). Major comparison types are (1) case versus healthy control, (2) sex-specific DE, (3) genotype-driven DE (i.e., apolipoprotein E (*APOE*) ε4/ε4 vs *APOE* ε3/ε3; TREM2^{R47H} vs common variants), and (4) others. Each comparison was accompanied by network analysis to reveal PPI modules, functional analyses to reveal the enriched pathways and biological processes, and gene set enrichment analyses for target and drug screening for more than 700,000 chemical perturbation profiles. We performed an exhaustive search of cell-cell interactions (CCIs) using a comprehensive ligand-receptor interaction (LRI) network that we have compiled (Table S4), achieving over 400 CCI analyses.
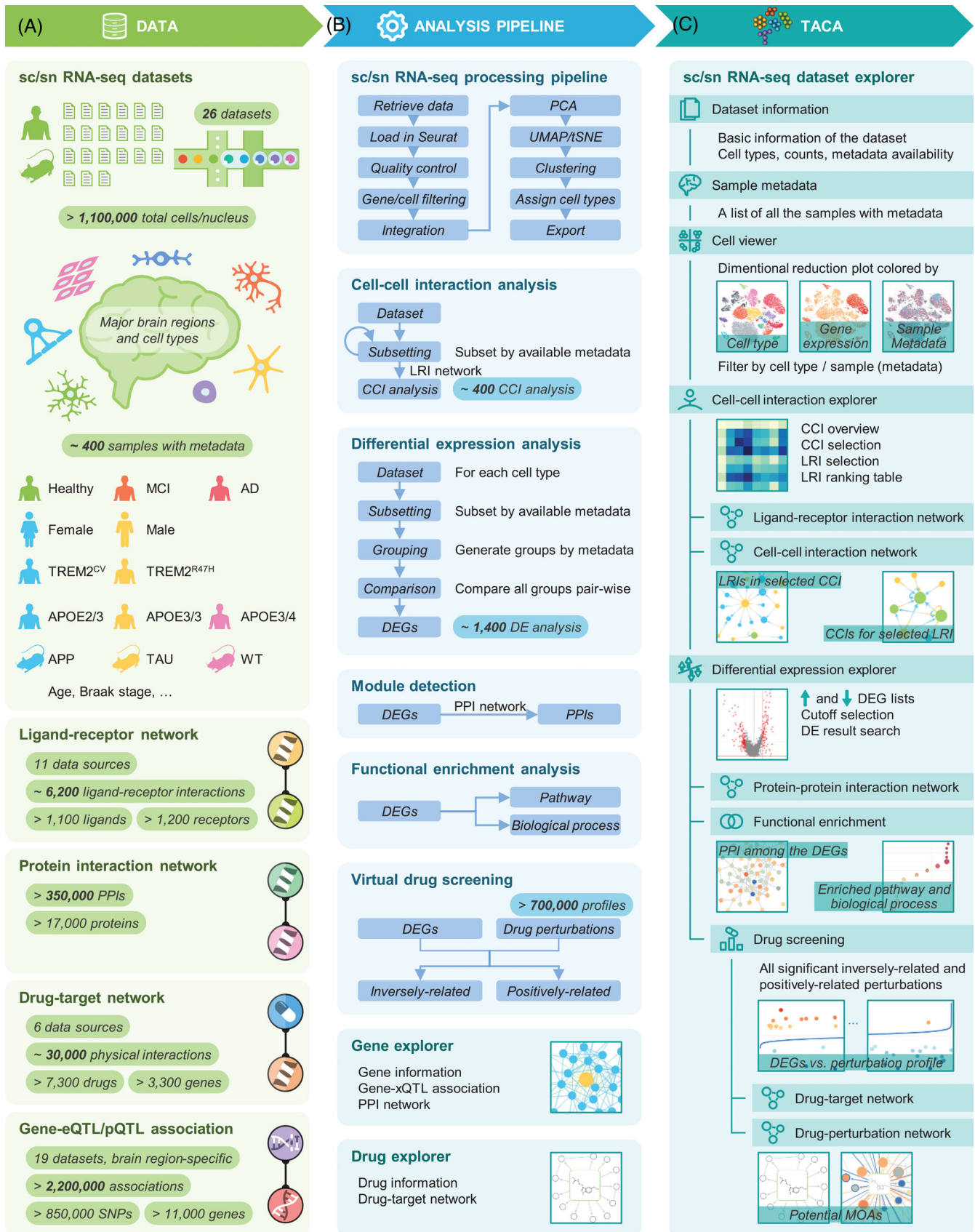
FIGURE 1 Overview of the information architecture and functions of The Alzheimer's Cell Atlas. (A) We have collected and assembled multiple types of data and networks, including single-cell/nucleus (sc/sn) RNA-sequencing (RNA-seq) data sets, ligand-receptor interactions (LRIs), protein-protein interactions (PPIs), drug-target interactions, and gene-quantitative trait locus (QTL) associations. (See Table S1 and Supplementary Methods for more details of the data sources and preprocessing steps.) In total, we obtained over 1.1 million cells/nuclei from the

All results were integrated into a new web service, TACA, with interactive data set, gene, and drug explorers, and a wide range of visualization tools such as dimensional reduction plot for cell types and gene expressions, volcano plot, and PPI network for the differentially expressed genes (DEGs), LRI network for CCIs, and mechanism-of-action plot for chemical perturbation profiles against the DEGs. All visualized networks can be modified interactively, downloaded as images, and exported for use on the users' own computers. All other types of visualization tools provide panning, scaling, selecting, and downloading as images, and offers informative messages when data points are hovered on.

3.2 | Interface and main functions of TACA

On the home page of TACA (Figure 2A), users can search for genes and drugs, which will lead users to their respective explorer pages. In the gene explorer (Figure 2B), the basic information, gene-xQTL associations (including expression quantitative trait locus [eQTL] and protein quantitative trait locus [pQTL]) (Table S5), and PPI network centered with the selected gene are shown. Gene-xQTL associations are categorized as positive ($\beta > 0$) and negative ($\beta < 0$) associations in two separate tables. The PPI network can help identify important neighbor genes (blue nodes) that may serve as targets of drugs to indirectly affect the gene of interest (yellow node). In the drug explorer (Figure 2C), basic drug information and the drug-target network of the selected drug are shown.

The home page lists all the sc/sn data sets and serves as their entry points. Each data set is shown in a dedicated data set explorer page. The data set explorer (Figure 2D) is composed of a navigation panel on the left (Figure 3A,B) and a shared space on the right for the currently selected tool from the navigation panel (Figure 3C–G). The tools in the navigation panel are organized in a tree format corresponding to the analysis pipeline. For example, DE is the upstream analysis of drug screening (downstream analysis) that utilized the DEGs, whereas drug screening is the upstream analysis of drug-perturbation network (downstream analysis) (Figure 3A). The downstream tool buttons are grayed out initially, and are (re)enabled when an upstream analysis is selected. Selecting a different upstream analysis will reset its

downstream tools. For example, when a DE comparison is selected, its associated PPI network, functional enrichment analysis, and drug screening results may become available to view using the buttons on the DE tool page initially and can be accessed later from the navigation panel until another DE selection is made. The details of the currently selected DE comparison are shown below the DE navigation button, similar for the drug screening and CCI analysis tools.

In the navigation panel, the first two buttons provide access to basic data set information and a table for sample metadata (Figure 3C). "Cell Viewer" offers a versatile dimensional reduction plot that has three coloring modes (Figure 3D), by cell types (i.e., microglia), by sample identities or sample metadata fields (e.g., TREM2 variants), and by gene expressions (e.g., *APOE*) in which cells are colored by a gradient. It is notable that "cell viewer" comes with a full-featured sample and/or cell type selector. For samples, users can choose to show or hide each sample individually, or by filtering all samples with one or more metadata fields (e.g., selecting all male mild cognitive impairment [MCI] samples).

In the "differential expression" tool, all DE comparisons can be found by selecting "strategy," "comparison," and "cell type." Once selected, the DE comparison's description, volcano plot, number of DEGs, and downstream analysis availabilities are shown, together with two tables at the page bottom for up- and down-expressed genes (Figure 3E). The gene names of the top DEGs with smallest false discovery rates (FDRs) are shown in the volcano plot, and are hidden upon clicking. In the two data tables, users can click the genes to open corresponding gene explorer pages. In TACA, we predefined nine sets of DEG cutoffs using fold change (FC) and FDR. These cutoffs can be selected in the table that shows the number of DEGs. The initial access points to the downstream tools are three buttons also found on this page.

In the "drug screen" tool (Figure 3F), all significant inversely and positively related perturbations are shown in two separate tables. For readability we display the compound name instead of the IDs of the perturbations (referred to as "signature" ID in Connectivity Map [CMap] L1000) in the tables. Once a perturbation is selected, its details are shown below, with two buttons for opening the drug target and perturbation network tools. These networks along with other ones are explained in the next section. At the bottom of the page is a scatter/line hybrid plot that visualizes the relationship between the perturbation profile and the selected DE comparison. The perturbation profile is

transcriptomic data sets. We curated the metadata of the samples in the data sets from the GEO database and original publications, which enabled a comprehensive analysis of differential expression (DE) comparisons and cell-cell interactions (CCIs). AD, Alzheimer's disease; CV, common variant; MCI, mild cognitive impairment. (B) The analysis pipeline of TACA. We adopted a standard sc/sn RNA-seq processing pipeline as shown. We referred to the original publications of these data sets for cutoffs for gene and cell filtering, dimensional reduction technique selection (i.e., UMAP or tSNE), marker genes for detecting cell types, and other additional processing steps if used in the original publication. Otherwise, we integrated the quality-controlled (cells filtered by mitochondria gene expression and number of features detected, etc.) data sets, performed dimensional reduction and clustering to annotate cell types, and exported the processed data for use in downstream analyses and the TACA webserver. For DE and CCI analyses, we defined possible analysis strategies and our pipeline conducted these analyses systematically (see Supplementary Methods). The differentially expressed genes (DEGs) were analyzed subsequently for PPIs, functional enrichment analysis, and virtual drug screening against over 700,000 chemical perturbation profiles. (C) Overview of the main tools (indicated by the tabs) and visualization types (indicated by the sample charts) available in the gene, drug, and sc/sn RNA-seq data set explorers in TACA. The tools in the data set explorer are organized as trees corresponding to the analysis pipeline. For example, drug-screening results can be accessed from the DE tool, when a specific DE comparison is selected. TACA incorporated several types of network visualizations for various types of biological relationships. These tools and visualizations are explained in more detail in Figures 2–4 and the Results section. MOA, mechanism of action.
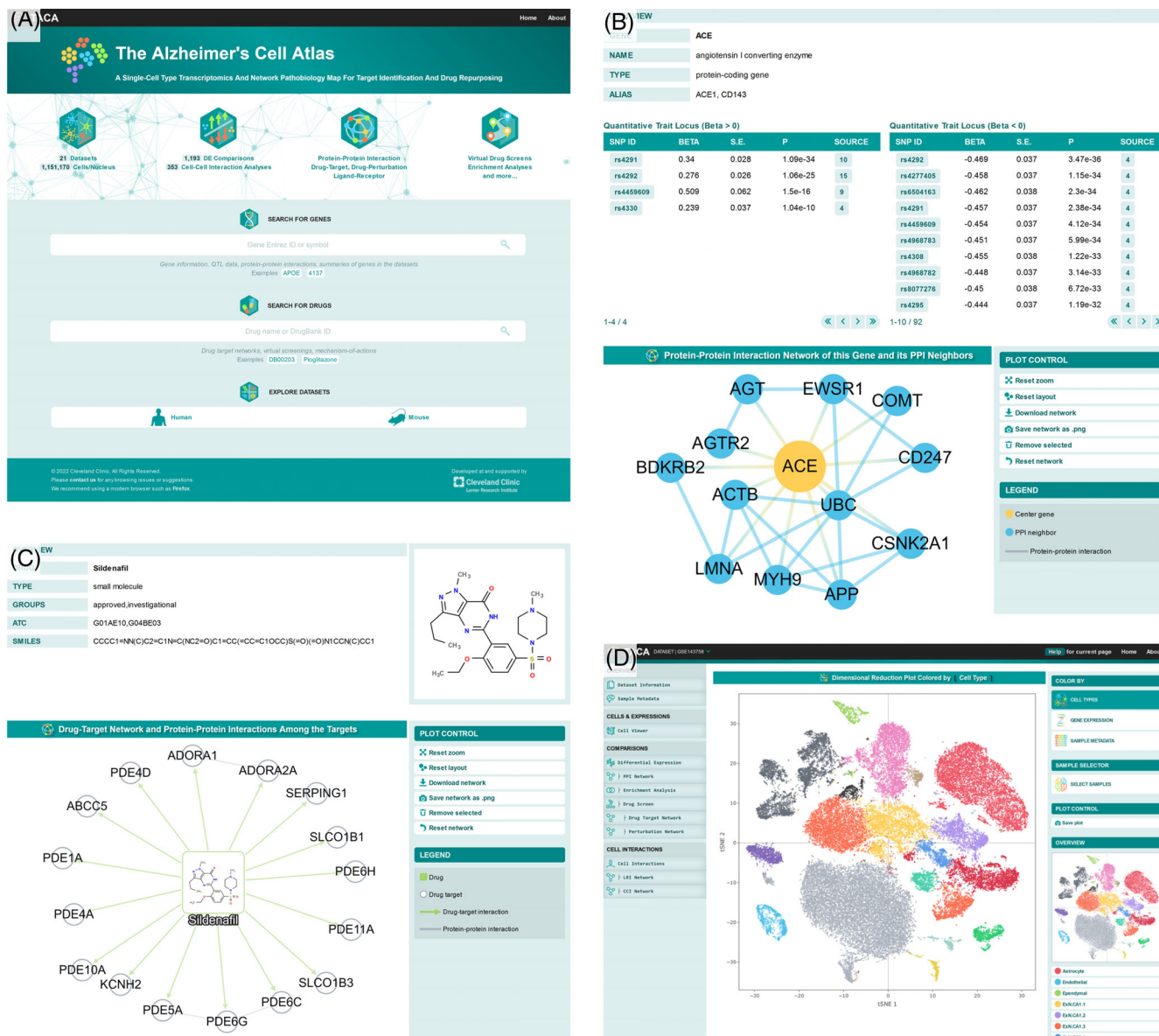
FIGURE 2 Drug, gene/target, and data set explorers in TACA. (A) The home page provides search tools for genes (B) and drugs (C) that direct users to the gene and drug explorers. All data sets in TACA can be listed by clicking the "human" or "mouse" buttons, and each data set has its own data set explorer page (D). (B) A gene explorer page shows the basic gene information, gene-quantitative trait locus (QTL) associations, and protein-protein interaction (PPI) neighbors of the gene of interest. (C) A drug explorer page shows the basic drug information, the structure, and the drug-target network of the drug of interest. PPIs among the targets are shown as gray edges. (D) A data set explorer page that currently displays the dimensional reduction plot colored by cell types. Various tools can be accessed from the navigation panel on the left side of the page. Several tools are grayed out upon page loading, indicating that they are downstream analyses whose results become available to view only when the upstream analysis is selected. The help information for each tool can be accessed using the "help" button in the top header.

shown as a blue line, in which genes (x-axis) are always in ascending order by their Z scores (y-axis) in the profile. The DEGs (dots) are x-positioned according to the genes in the perturbation profile, and are y-positioned and colored by their $\log_2$FC. For inversely related perturbations, the up-DEGs (warm color, above x-axis) tend to locate to the left, indicating that they are downregulated by the perturbation, and the down-DEGs (cold color, below x-axis) tend to locate to the right, indicating they are upregulated by the perturbation. This pattern is reversed for the positively related perturbations.

In the "cell interactions" tool (Figure 3G), all CCI analyses are found by selecting "strategy" and "analysis." Once a CCI analysis is selected, a heatmap is shown for the number of significant LRIs in all cell type pairs. The grids in the heatmap can be selected, and the significant LRIs for the selected cell-cell pair are populated in a table below. In another table, all LRIs that are significant in at least one cell-cell pair are listed in descending order by the number of significant cell-cell pairs. Two network visualizations can be accessed from this tool page for these two tables, respectively.
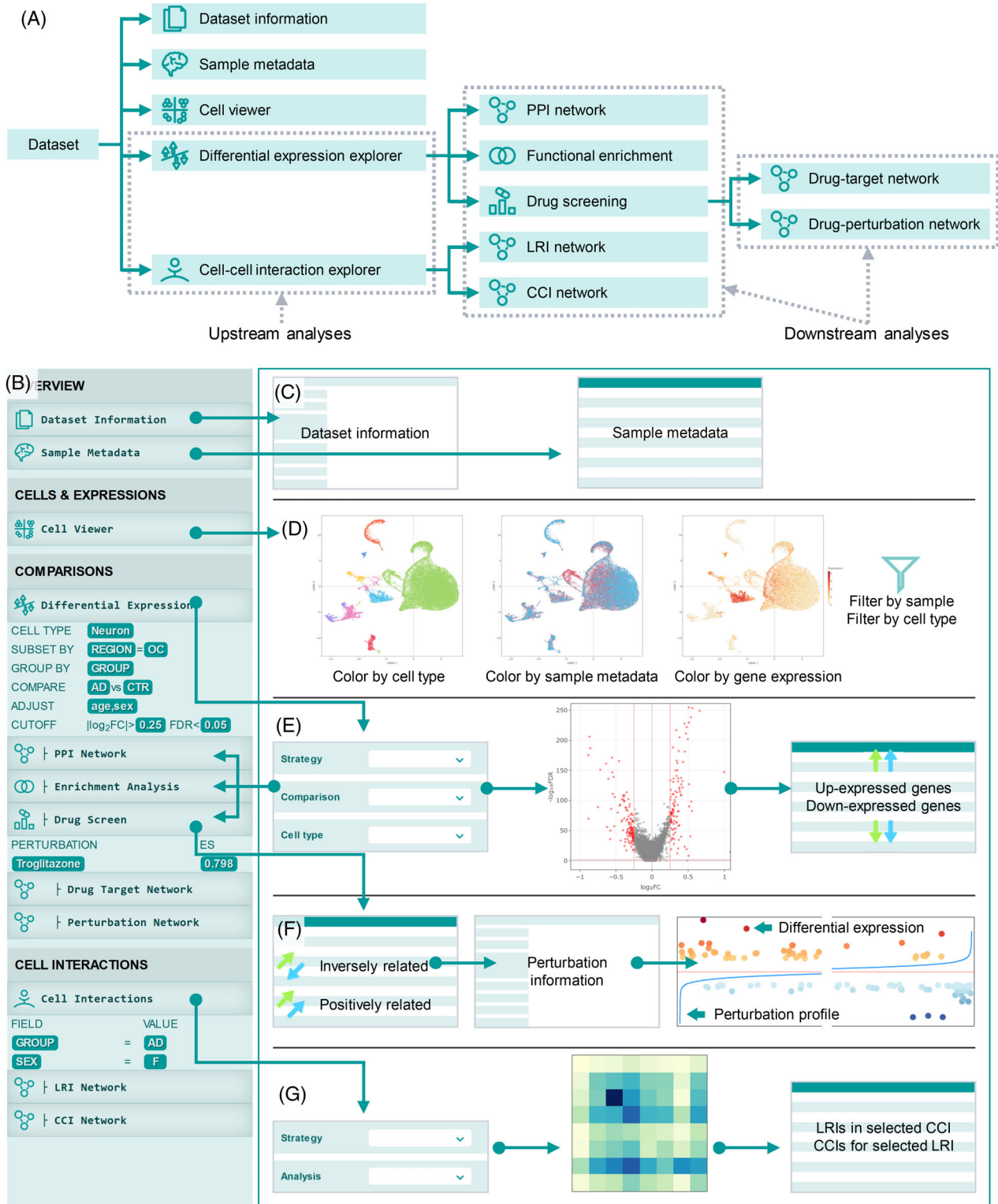
FIGURE 3 The data set page in TACA. (A) The main functions of the data set page are organized corresponding to the analysis pipeline. (B) A closer view of the navigation panel. Information regarding the currently selected analyses is shown in the navigation panel. (C) Basic data set information and a list of samples with metadata can be accessed by corresponding buttons. (D) In "cell viewer," the dimensional reduction plot has three coloring modes, by cell types, by gene expression (color gradient), and by samples or sample metadata. A full-featured cell selection tool based on cell type or sample (metadata) is offered. (E) In "differential expression" tool, all analyses are organized as "strategy," "comparison," and "cell type" (see Supplementary Methods). Once selected, a volcano plot of the comparison is shown, with those significantly differentially expressed genes colored in red. Two tables show the up- and down-expressed genes, respectively. The downstream analyses can be accessed from this tool (indicated by arrows pointing back to the navigation panel). (F) Drug-screening results are categorized as either inversely related (i.e., the perturbation leads to opposite gene expression pattern to that of the selected DE comparison) or positively related (i.e., the perturbation leads to

3.3 | Drug/gene/cell network visualizations in TACA

TACA offers seven types of network visualization tools, among which five (Figure 4) are found in the data set explorer.

In the "differential expression" tool, when a DE is selected, the "PPI Network" becomes accessible that shows the PPIs among the top 200 DEGs with the smallest FDRs (Figure 4A). Node colors and sizes indicate $\log_2 FC$ and FDR, respectively.

In the "drug screen" tool, when a perturbation is selected, its "drug target network" and "perturbation network" may become accessible. In the "drug target network" (Figure 4B), the targets of the drug are shown with the PPIs among them. Targets are colored by $\log_2 FC$ if they are also DEGs. This network shows the DEGs from the selected DE comparison that can be targeted directly by the selected drug, or targeted indirectly through PPIs with the drug's targets. In the "perturbation network" (Figure 4C), the inverse or positive relations of the DE and perturbation are visualized. Figure 4C shows an example of inverse relation, in which the up-DEGs (warm-colored nodes) are downregulated by the perturbation (cold-colored edges and borders), whereas the down-DEGs (cold-colored nodes) are upregulated by the perturbation (warm-colored edges and borders).

In the "cell interactions" tool, when a CCI is selected, "LRI Network" becomes available when a specific pair of cell types is selected from the heatmap (Figure 4D), and "CCI Network" becomes available when a specific LRI is selected from the table (Figure 4E). "LRI Network" shows the significant LRIs in the selected pair of cell types. Ligands and receptors are denoted by different colors. In "CCI Network," cell types are displayed instead, showing the cell types hosting the ligand that interact with cell types hosting the receptor. For example, using the data set (GEO ID: GSE98969) that led to the original discovery of DAM,[13] we found that the APOE-TREM2 interaction was one of the top significant LRIs in multiple CCIs (Figure 4E) in the 5XFAD mouse, including DAM-DAM and DAM-microglia (Figure 4D). This observation is consistent with those of previous studies that demonstrated the important roles of APOE-TREM2 interaction in modulating phagocytosis and mediating the transition from homeostatic microglia to DAM.[13,33–35]

3.4 | Discovery of repurposable drugs for AD using TACA

In this example, we selected data set "GSE148822." In the "differential expression" tool, we selected the strategy 'SUBSET by "REGION" – GROUP by "GROUP" – ADJUST by "AGE,SEX"', comparison 'SUBSET = "OC" – COMPARE GROUPs "AD" versus "CTR,"' and cell type "Neuron." In other words, here we are exploring the DE results of comparing occipital cortex (OC) samples in AD patients versus those in non-demented controls (CTR) for the cell type neuron. The DE comparison resulted in 79 DEGs, such as SLC1A3, SLC1A2, SPRED1, GPC5, MBP, and DDX24. It has been reported that members from the solute carriers (SLCs) family may be associated with neurodegenerative diseases.[36] SPRED1 may be involved in tauopathy.[37] By clicking "drug screen" below the "Number of DEGs" table, the page is switched to the "drug screen" tool. As the comparison is AD versus CTR, the desired relationship is, therefore, "inversely-related" (such that up-DEGs in AD are downregulated by the drug perturbation and the down-DEGs in AD are upregulated by the drug perturbation to achieve an "rescued" effect). In this table, one perturbation (troglitazone) has a significant enrichment score. By clicking this perturbation, we see that most of the up-DEGs are downregulated in the perturbation profile, and most of the down-DEGs are upregulated by the drug (Figure 5A,B). By comparing the strongly perturbed genes (e.g., *STAT1*, *CLU*, *GPM6A*, *CST3*, *SLC1A2*) (Figure 5A,B) with a list of AD-associated risk genes that were compiled in a previous study,[10,12] we found that clusterin (*CLU*),[38] which is significantly up-expressed ($\log_2 FC = 0.453$, FDR = 0.000006) in AD versus CTR, is strongly downregulated by troglitazone (Z score = −1.767). We found that one of troglitazone's physical interacting targets (Figure 5C), transient receptor potential cation channel subfamily M member 3 (TRPM3), is significantly overexpressed ($\log_2 FC = 0.977$, FDR = 0.0002) in AD versus CTR. Troglitazone is a TRPM3 inhibitor ($IC_{50} = 12 \mu M$).[39] It also downregulated TRPM3 in this perturbation (Z score = -0.562). These results suggest that troglitazone may have a beneficial effect for AD neurons by reducing the levels of two up-expressed genes in AD. It is possible that other inversely perturbed genes in Figure 5A and B can explain the beneficial effect.

3.5 | Discovery of potential pathobiology of AD using TACA

Here we show a case of how we identify potential pathobiology of AD in cell type–specific manners using TACA. Previous studies have shown that the transcription factor EB (TFEB) may have a protective role against AD because the upregulation of TFEB alleviated AD pathologies in mice and cells.[40] TFEB is a master regulator of lysosomal biogenesis and plays important roles in autophagy and mitophagy,[40,41] which were shown to be associated with AD pathology.[42,43] Here, we show that by using three data sets (GSE147528_EC, GSE147528_SFG, and GSE148822) from two studies,[15,26] we found that TFEB was significantly downregulated when we compared AD patients with healthy or less-severe AD patients (Figure 6). The cell dimensional reduction

similar gene expression pattern to that of the selected DE comparison). Gene expression patterns in both the DE comparison (dots whose colors and y positions indicate expression fold change) and the selected chemical perturbation (blue line for gene expressions in ascending order) are plotted. (G) Cell-cell interactions (CCIs) (i.e., cellular communications among cell types) analyses are organized similar to that of the "differential expression" tool. Once an analysis is selected, the numbers of significant ligand-receptor interactions (LRIs) in all cell type pairs are visualized in a heatmap. The results are organized in two tables, showing the significant LRIs in the selected CCI and the number of CCIs in which a certain LRI is significant, respectively.
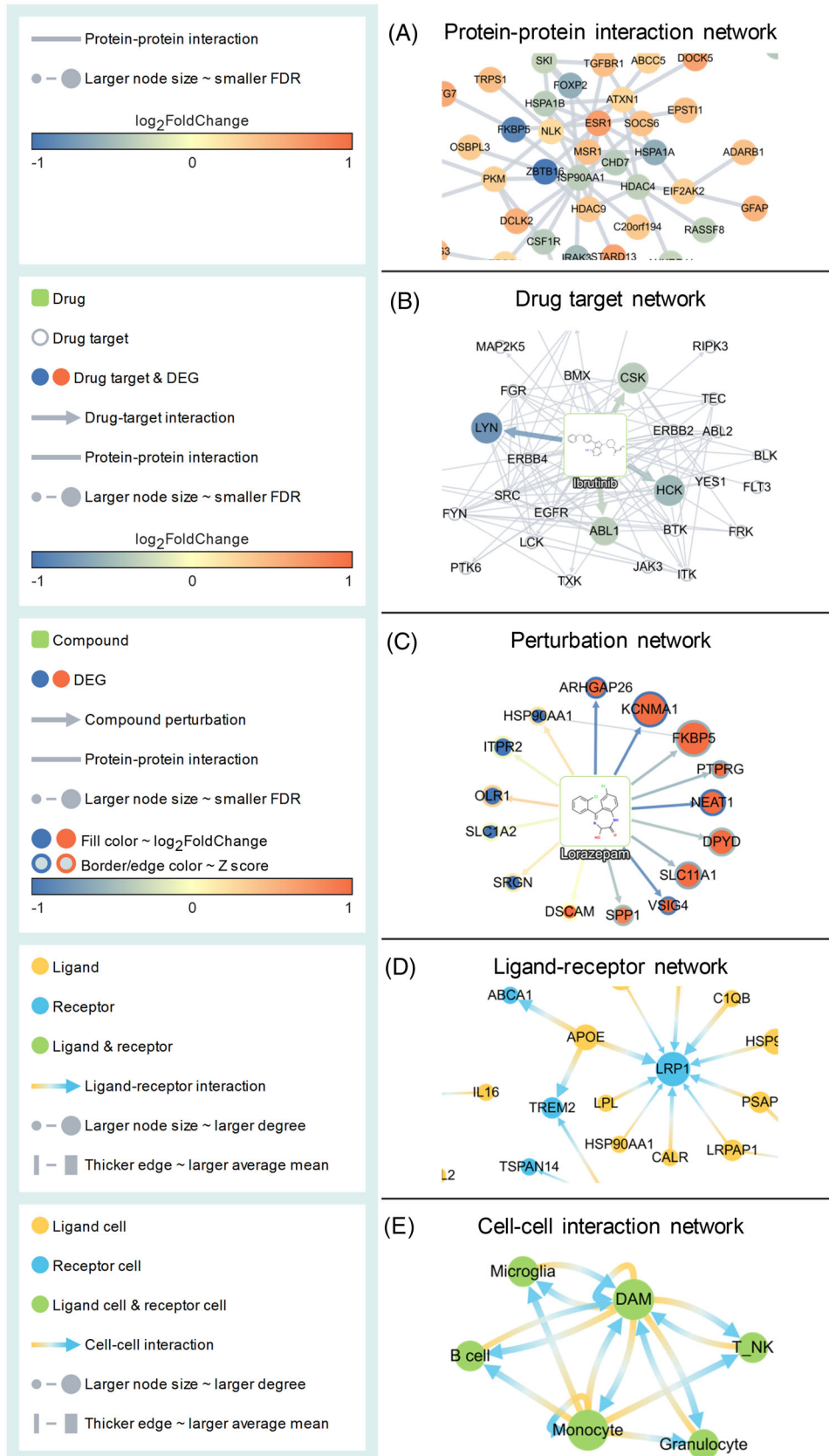
FIGURE 4 The network visualizations in TACA. (A) Protein-protein interaction (PPI) network of the differentially expressed genes. Node colors indicate $\log_2$ fold change ($\log_2$FC), and node sizes indicate false discovery rate (FDR). (B) Drug target network of the selected drug. Differentially expressed drug targets are colored by $\log_2$FC. PPIs among the targets are shown. (C) Perturbation network that visualized the inverse relation or positive relation of the differential expression results and gene profiles of a chemical perturbation. A maximum of 50

plots, gene expression plots (Figure 6A–C), and DE analyses results (Figure 6D–F) can be found in TACA as explained in previous sections.

In GSE147528_EC and GSE147528_SFG (Figure 6A,B), we found that as Braak stage increases, the expression of TFEB significantly decreases ($|\log_2 FC| > 0.25$ and FDR < 0.05) in both the entorhinal cortex (EC) region and superior frontal gyrus (SFG) region from post-mortem brain tissue of male donors (Figure 6D,E). In GSE148822, we found that the expression of TFEB was inversely associated with AD disease progression using male samples from OC and occipitotemporal cortex (OTC) regions (Figure 6C,F). However, this effect is not observed in female patients ($|\log_2 FC| < 0.25$ or FDR > 0.05). In addition, TFEB is highly expressed in the oligodendrocytes (Figure 6A–C), consistent with results of a previous study that TFEB plays important roles in myelination in the oligodendrocytes.[44] These observations illustrate that TACA offers a useful tool for identifying potential pathobiology of AD in cell type–specific manners.

4 | DISCUSSION

We present TACA, a web portal and database with strong potential for the identification of cell type–specific AD pathobiology as well as target discovery for drug repurposing. We collected and processed a large amount of data, including sc/sn RNA-seq transcriptomic data sets and many types of networks. Our first version of TACA achieved over 1.1 million cells/nuclei and ≈1400 differential expression and 400 cell-cell interaction analyses with various downstream analyses. We will continue to expand TACA by adding new sc/sn RNA-seq data sets and new types of visualizations and analyses.

TACA offers a highly organized and interactive interface. Currently, there are 12 types of visualization tools throughout the data set, gene, and drug explorers. TACA's many types of network visualizations will play important roles in showing PPIs among DEGs, understanding cell type communications by LRIs, and revealing potential mechanisms of action of chemical perturbations against the DE comparisons and so on.

As examples, we used the data and tools provided in TACA, and identified that troglitazone may have a protective effect for AD neurons. We found that it can lower the expression levels of CLU (known AD risk–associated gene)[38] and TRPM3 (direct target of troglitazone) that are both significant up-DEGs in AD neurons. Previous studies have reported that troglitazone has a protective effect on neurodegenerative disorders, such as AD.[45] Yet, the underlying molecular mechanisms are not fully understood. A potential explanation is that inhibition of cyclin-dependent kinase 5 (CDK5) activity by troglitazone repressed tau-Thr231 phosphorylation.[45] Our case study shows that the virtual drug screening in TACA discovered troglitazone for AD without this prior knowledge, and identified two additional potential mechanisms

of action for the beneficial effect. Our second case study of TFEB shows that TACA can be validated at mechanistic level, and we further found a male-specific protective effect of TFEB.

We envision that TACA will be a highly valuable resource for both basic and translational research in AD, as it provides abundant information for AD pathobiology and actionable systems biology tools for therapeutic discovery. Our framework can guide future AD sc/sn analyses and cell type–specific pathobiology and target discovery by providing numerous examples of data processing, analysis, and interpretation. Moreover, our framework can be broadly applied to other diseases. TACA will be regularly updated to include up-to-date sc/sn RNA-seq AD data sets.

## 4.1 | Collaborative interactions with other sc/sn RNA-seq and AD resources

To date, several useful bioinformatics tools have been developed for a broader range of sc/sn data set exploration, such as Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) and CELLxGENE (https://cellxgene.cziscience.com/), and for AD studies, such as Agora (https://agora.adknowledgeportal.org/genes) and the Alzheimer's Disease Atlas (https://adatlas.helmholtz-muenchen.de/).[46] We envision that it would be beneficial to the AD research community if TACA could establish collaborative work with these resources in the future. For example, a pipeline may be implemented to automatically import the annotated AD sc/sn data sets from Single Cell Portal, and our analysis pipeline will conduct analyses such as DE, CCI, and drug screening. The analyses outputs can be integrated into (or linked from) tools such as Agora for a more comprehensive view of genes and networks in a cell type–specific manner for rapid data sharing.

4.2 | Limitations and future directions

We acknowledge several limitations. First, although we included 26 AD data sets, more data sets have become available during the development of TACA. We will expand TACA in the following directions. (1) We will continue to process the sc/sn RNA-seq data sets as we did for the first phase of the data sets in TACA, as well as allowing user-supplied processed data sets in .rds format to be added using a pipeline that we have developed for this purpose. (2) We will focus on adding data sets from more diverse populations (e.g., African American, Asian populations,[47] and other minority populations), brain regions, and other AD tissue types (e.g., peripheral blood mononuclear cell [PBMC] and cerebrospinal fluid [CSF]). (3) We will integrate other types of omics data, such as scATAC-seq, and offer multi-omics inte-

---

differentially expressed genes (DEGs) with the lowest Z scores and 50 with the highest Z scores in the perturbation profile are shown in the network. Gene nodes are colored and sized by their $\log_2 FC$ and FDR, respectively, whereas their border colors and edge (to the compound) colors indicate the Z scores in the perturbation profile. As a result, plot for inversely related perturbation and DEGs will have inverse node and edge color, whereas positively related perturbation and DEGs will have similar node and edge color. PPIs among the DEGs are shown as gray edges. (D) Ligand-receptor interaction (LRI) network for the selected cell-cell interaction. (E) Cell-cell interaction network for the selected LRI.
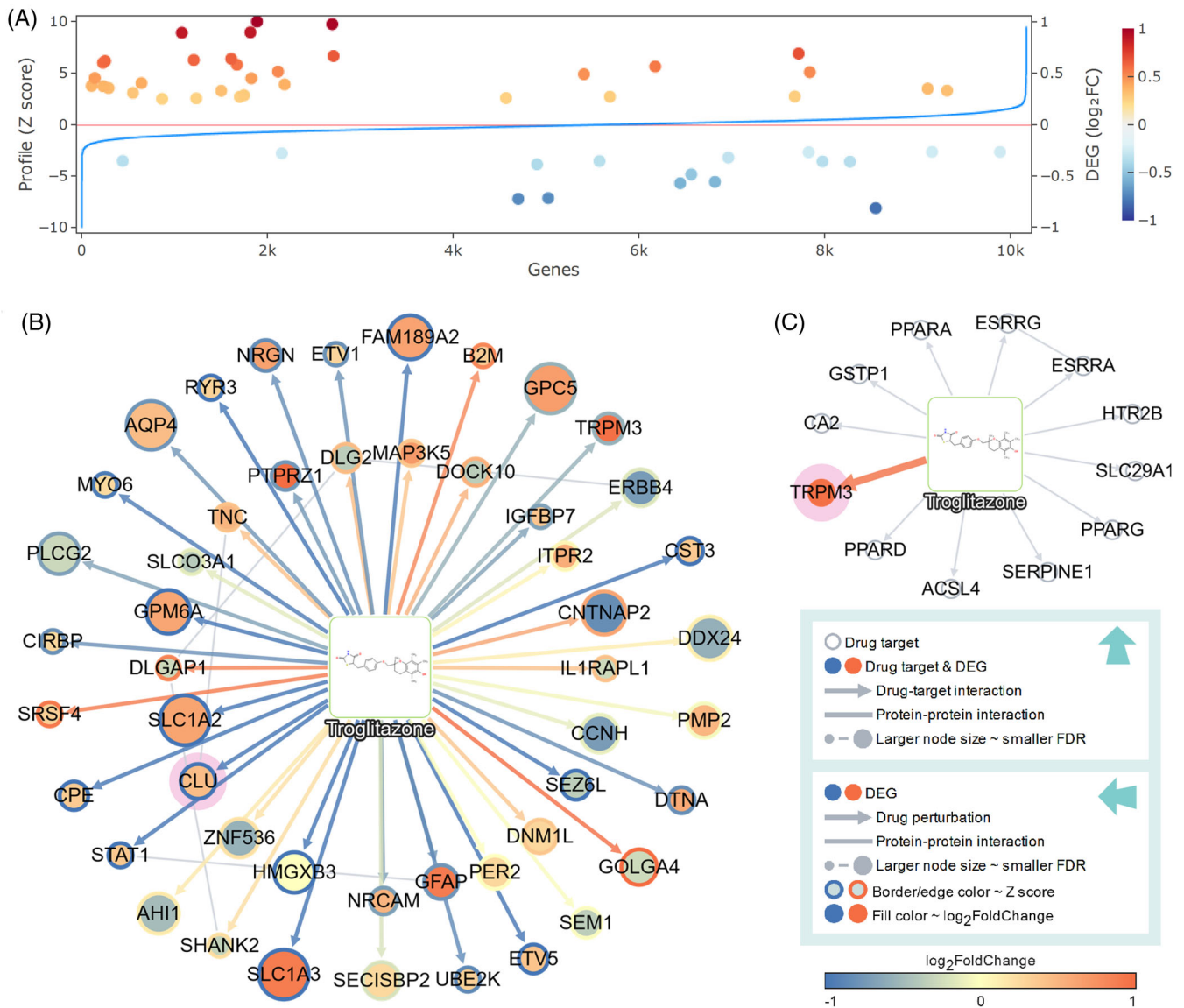
FIGURE 5 Case study: single-cell transcriptomics-based drug screening. (A) This plot shows the inverse relationship between the selected drug perturbation (blue line, genes ordered in ascending order by their expression Z scores) and the differential expression (DE) (colored dots, x-positioned according to the perturbation profile) profiles. The up-differentially expressed genes (DEGs) (warm color) are downregulated by the perturbation, whereas the down-DEGs (cold color) are upregulated by the perturbation. (B) A drug perturbation network that shows the (a maximum of) 50 DEGs with the lowest Z scores and 50 with the highest Z scores in the perturbation profile are shown in the network. In inversely related drug perturbation and DE profiles, the node color (indicate DE profile) and edge/border color (indicate drug perturbation profile) of the majority of the nodes are shown in opposite colors. (C) A drug target network colored by the DEG profiles. Non-DEG targets are shown as gray circles.

gration analyses.[48,49] We will add additional tables on the gene page to show other omic layers, such as proteomic and metabolomic data from the AD knowledge portal and The Alzheimer's Disease Metabolomics Consortium (ADMC).[50] (4) We will expand TACA for other neurodegenerative diseases, such as Parkinson disease (PD) and amyotrophic lateral sclerosis (ALS). Second, although we have curated the metadata from the GEO database and original publications, the availability of metadata varies among the data sets, and those with limited metadata have, therefore, limited DE comparisons and CCI analysis results. We recommend that researchers make their sample metadata available as

complete as possible, since these metadata can significantly improve the reusability of the data sets. We will add more analysis results for existing data sets if these metadata become available. Third, although we integrated data from many sources to generate the human protein interactome, drug-target network, and ligand-receptor network, these networks are still incomplete and will be expanded. Fourth, the current "cell viewer" is optimized for showing large numbers of cells, but a "subset" function that loads only a subsetted data set with fewer cells may be useful to accelerate the performance on older generation computers. Fifth, we predefined nine sets of DE cutoffs for generating DEGs for
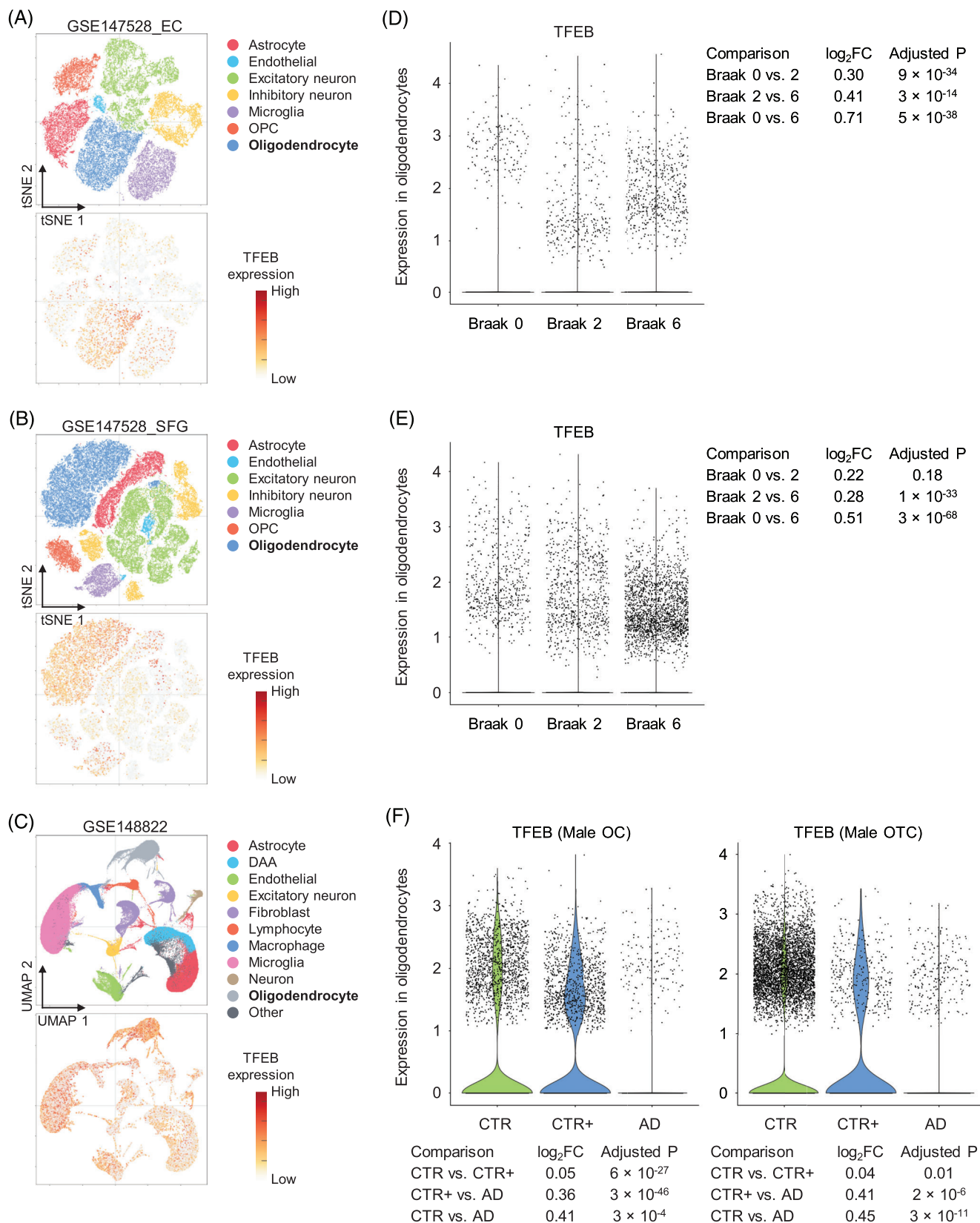
FIGURE 6 Case study: discovery of potential pathobiology of AD using TACA. (A–C) Dimensional reduction plots and expression plots for transcription factor EB (TFEB) from three data sets in TACA. (D–F) TFEB was significantly downregulated when we compared AD patients with healthy or less-severe AD patients. CTR, non-demented controls; CTR+, non-demented controls with mild amyloid beta pathology; DAA, disease-associated astrocyte; EC, entorhinal cortex; FC, fold change; OC, occipital cortex; OPC, oligodendrocyte progenitor cell; OTC, occipitotemporal cortex; SFG, superior frontal gyrus.

downstream analyses, such that drug screening can be pre-calculated. In future updates, we will further improve the drug-screening computational efficiency to allow user-defined DE cutoffs. Finally, advanced artificial intelligence/machine-learning techniques, such as deep generative model and transfer-learning approaches, can be applied for sc/sn data integration (among the AD data sets or with non–disease-centric datasets such as the Tabula Sapiens[51]) and analysis to identify novel/rare cell types and states.[52]

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

Dr. Cummings has provided consultation to AB Science, Acadia, Alkahest, AlphaCognition, ALZPathFinder, Annovis, AriBio, Artery, Avanir, Biogen, Cerevel, Clinilabs, Cortexyme, Diadem, EIP Pharma, Eisai, Genentech, Green Valley, Grifols, Janssen, Karuna, Lexeo, Lilly, Lundbeck, LSP, Merck, NervGen, Novo Nordisk, Oligomerix, Otsuka, PharmatrophiX, PRODEO, Prothena, ReMYND, Resverlogix, Roche, Signant

Health, Suven, Unlearn AI, Vaxxinity pharmaceutical, assessment, and investment companies. Dr. Leverenz has received consulting fees from consulting fees from Vaxxinity, grant support from GE Healthcare, and serves on a Data Safety Monitoring Board for Eisai. The other authors have declared no competing interests. Author disclosures are available in the supporting information.

REFERENCES

1. Alzheimer's Association. 2022 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2022;18(4):700-789. doi: 10.1002/alz.12638
2. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther*. 2014;6(4):37. doi: 10.1186/alzrt269
3. Mielke MM, Vemuri P, Rocca WA. Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences. *Clin Epidemiol*. 2014;6:37-48. doi: 10.2147/CLEP.S37929
4. Rocca WA, Boyd CM, Grossardt BR, et al. Prevalence of multimorbidity in a geographically defined American population: patterns by age, sex, and race/ethnicity. *Mayo Clin Proc*. Oct 2014;89(10):1336-1349. doi: 10.1016/j.mayocp.2014.07.010
5. Podcasy JL, Epperson CN. Considering sex and gender in Alzheimer disease and other dementias. *Dialogues Clin Neurosci*. 2016;18(4):437-446.
6. Nebel RA, Aggarwal NT, Barnes LL, et al. Understanding the impact of sex and gender in Alzheimer's disease: a call to action. *Alzheimers Dement*. 2018;14(9):1171-1183. doi: 10.1016/j.jalz.2018.04.008
7. Tasaki S, Gaiteri C, Mostafavi S, De Jager PL, Bennett DA. The molecular and neuropathological consequences of genetic risk for Alzheimer's dementia. *Front Neurosci*. 2018;12:699. doi: 10.3389/fnins.2018.00699
8. Khramtsova EA, Davis LK, Stranger BE. The role of sex in the genomics of human complex traits. *Nat Rev Genet*. 2019;20(3):173-190. doi: 10.1038/s41576-018-0083-1
9. Fang J, Pieper AA, Nussinov R, et al. Harnessing endophenotypes and network medicine for Alzheimer's drug repurposing. *Med Res Rev*. 2020;40(6):2386-2426. doi: 10.1002/med.21709
10. Fang J, Zhang P, Zhou Y, et al. Endophenotype-based in silico network medicine discovery combined with insurance record data mining identifies sildenafil as a candidate drug for Alzheimer's disease. *Nature Aging*. 2021;1(12):1175-1188. doi: 10.1038/s43587-021-00138-z
11. Fang J, Zhang P, Wang Q, et al. Artificial intelligence framework identifies candidate targets for drug repurposing in Alzheimer's disease. *Alzheimers Res Ther*. 2022;14(1):7. doi: 10.1186/s13195-021-00951-z
12. Zhou Y, Fang J, Bekris LM, et al. AlzGPS: a genome-wide positioning systems platform to catalyze multi-omics for Alzheimer's drug discovery. *Alzheimers Res Ther*. 2021;13(1):24. doi: 10.1186/s13195-020-00760-w
13. Keren-Shaul H, Spinrad A, Weiner A, et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell*. 2017;169(7):1276-1290.e17. doi: 10.1016/j.cell.2017.05.018
14. Habib N, McCabe C, Medina S, et al. Disease-associated astrocytes in Alzheimer's disease and aging. *Nat Neurosci*. 2020;23(6):701-706. doi: 10.1038/s41593-020-0624-8
15. Gerrits E, Brouwer N, Kooistra SM, et al. Distinct amyloid-beta and tau-associated microglia profiles in Alzheimer's disease. *Acta Neuropathol*. 2021;141(5):681-696. doi: 10.1007/s00401-021-02263-w
16. Ferreira D, Wahlund LO, Westman E. The heterogeneity within Alzheimer's disease. *Aging (Albany NY)*. 2018;10(11):3058-3060. doi: 10.18632/aging.101638
17. Xu J, Zhang P, Huang Y, et al. Multimodal single-cell/nucleus RNA sequencing data analysis uncovers molecular networks between disease-associated microglia and astrocytes with implications for drug

repurposing in Alzheimer's disease. *Genome Res*. 2021;31(10):1900-1912. doi: 10.1101/gr.272484.120

18. Sala Frigerio C, Wolfs L, Fattorelli N, et al. The major risk factors for Alzheimer's disease: age, sex, and genes modulate the microglia response to Abeta Plaques. *Cell Rep*. 2019;27(4):1293-1306 e6. doi: 10.1016/j.celrep.2019.03.099

19. Grubman A, Chew G, Ouyang JF, et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci*. 2019;22(12):2087-2097. doi: 10.1038/s41593-019-0539-4

20. Guttikonda SR, Sikkema L, Tchieu J, et al. Fully defined human pluripotent stem cell-derived microglia and tri-culture system model C3 production in Alzheimer's disease. *Nat Neurosci*. 2021;24(3):343-354. doi: 10.1038/s41593-020-00796-z

21. Zhou Y, Song WM, Andhey PS, et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med*. 2020;26(1):131-142. doi: 10.1038/s41591-019-0695-9

22. Zhong S, Wang M, Zhan Y, et al. Single-nucleus RNA sequencing reveals transcriptional changes of hippocampal neurons in APP23 mouse model of Alzheimer's disease. *Biosci Biotechnol Biochem*. 2020;84(5):919-926. doi: 10.1080/09168451.2020.1714420

23. Sierksma A, Lu A, Mancuso R, et al. Novel Alzheimer risk genes determine the microglia response to amyloid-beta but not to TAU pathology. *EMBO Mol Med*. 2020;12(3):e10606. doi:10.15252/emmm.201910606

24. Perez MJ, Ivanyuk D, Panagiotakopoulou V, et al. Loss of function of the mitochondrial peptidase PITRM1 induces proteotoxic stress and Alzheimer's disease-like pathology in human cerebral organoids. *Mol Psychiatry*. 2021;26(10):5733-5750. doi: 10.1038/s41380-020-0807-4

25. Lau SF, Chen C, Fu WY, et al. IL-33-PU.1 transcriptome reprogramming drives functional state transition and clearance activity of microglia in Alzheimer's disease. *Cell Rep*. 2020;31(3):107530. doi: 10.1016/j.celrep.2020.107530

26. Leng K, Li E, Eser R, et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat Neurosci*. 2021;24(2):276-287. doi: 10.1038/s41593-020-00764-7

27. Wang S, Mustafa M, Yuede CM, et al. Anti-human TREM2 induces microglia proliferation and reduces pathology in an Alzheimer's disease model. *J Exp Med*. 2020;217(9):e20200785. doi: 10.1084/jem.20200785

28. Najm R, Zalocusky KA, Zilberter M, et al. In vivo chimeric Alzheimer's disease modeling of apolipoprotein E4 toxicity in human neurons. *Cell Rep*. 2020;32(4):107962. doi: 10.1016/j.celrep.2020.107962

29. Lau SF, Cao H, Fu AKY, Ip NY. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proc Natl Acad Sci U S A*. 2020;117(41):25800-25809. doi: 10.1073/pnas.2008762117

30. McQuade A, Kang YJ, Hasselmann J, et al. Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nat Commun*. 2020;11(1):5370. doi: 10.1038/s41467-020-19227-5

31. Otero-Garcia M, Xue Y-Q, Shakouri T, et al. Single-soma transcriptomics of tangle-bearing neurons in Alzheimer's disease reveals the signatures of tau-associated synaptic dysfunction. *bioRxiv*2020.05.11.088591. doi: 10.1101/2020.05.11.088591

32. Mathys H, Davila-Velderrain J, Peng Z, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570(7761):332-337. doi: 10.1038/s41586-019-1195-2

33. Krasemann S, Madore C, Cialic R, et al. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity*. 2017;47(3):566-581 e9. doi: 10.1016/j.immuni.2017.08.008

34. Atagi Y, Liu CC, Painter MM, et al. Apolipoprotein E is a ligand for triggering receptor expressed on myeloid cells 2 (TREM2). *J Biol Chem*. 2015;290(43):26043-50. doi: 10.1074/jbc.M115.679043

35. Yeh FL, Wang Y, Tom I, Gonzalez LC, Sheng M. TREM2 binds to apolipoproteins, including APOE and CLU/APOJ, and thereby facilitates uptake of amyloid-beta by microglia. *Neuron*. 2016;91(2):328-40. doi: 10.1016/j.neuron.2016.06.015

36. Hu C, Tao L, Cao X, Chen L. The solute carrier transporters and the brain: physiological and pharmacological implications. *Asian J Pharm Sci*. 2020;15(2):131-144. doi: 10.1016/j.ajps.2019.09.002

37. Johne C, Matenia D, Li XY, Timm T, Balusamy K, Mandelkow EM. Spred1 and TESK1 – two new interaction partners of the kinase MARKK/TAO1 that link the microtubule and actin cytoskeleton. *Mol Biol Cell*. 2008;19(4):1391-403. doi: 10.1091/mbc.E07-07-0730

38. Ebbert MTW, Boehme KL, Wadsworth ME, et al. Interaction between variants in CLU and MS4A4E modulates Alzheimer's disease risk. *Alzheimers Dement*. 2016;12(2):121-129. doi: 10.1016/j.jalz.2015.08.163

39. Majeed Y, Bahnasi Y, Seymour VA, et al. Rapid and contrasting effects of rosiglitazone on transient receptor potential TRPM3 and TRPC5 channels. *Mol Pharmacol*. 2011;79(6):1023-30. doi: 10.1124/mol.110.069922

40. Xu Y, Du S, Marsh JA, et al. TFEB regulates lysosomal exocytosis of tau and its loss of function exacerbates tau pathology and spreading. *Mol Psychiatry*. 2021;26(10):5925-5939. doi: 10.1038/s41380-020-0738-0

41. Aman Y, Schmauck-Medina T, Hansen M, et al. Autophagy in healthy aging and disease. *Nat Aging*. 2021;1(8):634-650. doi: 10.1038/s43587-021-00098-4

42. Fang EF, Hou Y, Palikaras K, et al. Mitophagy inhibits amyloid-beta and tau pathology and reverses cognitive deficits in models of Alzheimer's disease. *Nat Neurosci*. 2019;22(3):401-412. doi: 10.1038/s41593-018-0332-9

43. Kobro-Flatmoen A, Lagartos-Donate MJ, Aman Y, Edison P, Witter MP, Fang EF. Re-emphasizing early Alzheimer's disease pathology starting in select entorhinal neurons, with a special focus on mitophagy. *Ageing Res Rev*. 2021;67:101307. doi: 10.1016/j.arr.2021.101307

44. Sun LO, Mulinyawe SB, Collins HY, et al. Spatiotemporal control of CNS myelination by oligodendrocyte programmed cell death through the TFEB-PUMA Axis. *Cell*. 2018;175(7):1811-1826.e21. doi: 10.1016/j.cell.2018.10.044

45. Cho DH, Lee EJ, Kwon KJ, et al. Troglitazone, a thiazolidinedione, decreases tau phosphorylation through the inhibition of cyclin-dependent kinase 5 activity in SH-SY5Y neuroblastoma cells and primary neurons. *J Neurochem*. 2013;126(5):685-95. doi: 10.1111/jnc.12264

46. Wörheide MA, Krumsiek J, Nataf S, et al. An integrated molecular atlas of Alzheimer's disease. *medRxiv* 2021.09.14.21263565. doi: 10.1101/2021.09.14.21263565

47. Jiang Y, Zhou X, Ip FC, et al. Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging. *Alzheimers Dement*. 2022;18(1):88-102. doi: 10.1002/alz.12369

48. Corces MR, Shcherbina A, Kundu S, et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet*. 2020;52(11):1158-1168. doi: 10.1038/s41588-020-00721-x

49. Morabito S, Miyoshi E, Michael N, et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet*. 2021;53(8):1143-1155. doi: 10.1038/s41588-021-00894-z

50. Toledo JB, Arnold M, Kastenmuller G, et al. Metabolic network failures in Alzheimer's disease: a biochemical road map. *Alzheimers Dement*. 2017;13(9):965-984. doi: 10.1016/j.jalz.2017.01.020

51. Tabula Sapiens C, Jones RC, Karkanias J, et al. The tabula sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science.* 2022;376(6594):eabl4896. doi: 10.1126/science.abl4896

52. Sikkema L, Strobl D, Zappia L, et al. An integrated cell atlas of the human lung in health and disease. *bioRxiv* 2022.03.10.483747. doi: 10.1101/2022.03.10.483747

SUPPORTING INFORMATION

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