Title

A Systems Biology Approach to Precision Medicine for Alzheimer’s Disease: Cell-type-directed Network-correcting Therapeutics and Transcriptomic Profiling across Major Risk Factors

Author

Li, Yaqiao

Publication Date

2024

A Systems Biology Approach to Precision Medicine for Alzheimer's Disease: Cell-type-directed Network-Correcting Therapeutics and Transcriptomic Profiling across Disease Risk Factors

by

Yaqiao Li

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Marina Sirota ______________________________ Marina Sirota

925B61AB9C41499...

Chair

Signed by:

Yadong Huang ______________________________ Yadong Huang

Signed by:

[illegible]D9AA1471...

Michael Keiser ______________________________ Michael Keiser

4DF1BD06D670465...

____

____

Committee Members

ACKNOWLEDGEMENTS

To my mother, who instilled in me a deep will for the pursuit of knowledge and provided me with the opportunity to pursue higher education. Your memory remains a constant source of inspiration and motivation, and I am profoundly grateful for all that you had done for me. To Marilyn and Jerry Jacobson, whose unconditional love and unwavering support have been invaluable throughout my graduate studies. I am deeply fortunate to have you in my life, and I extend my heartfelt appreciations. To Dana Buntrock and LeRoy Howard, who have profoundly influenced my life by instilling in me a strong commitment to pursuing a career that is gratifying and impactful. Your unwavering support, forward-thinking guidance, and life wisdom have been instrumental at every stage of my journey, and for that, I am sincerely appreciative.

To my husband, Ouwei Wang, who has proven to be my perfect match, the most supportive companion throughout graduate school, and a role model. I admire you for all that you have accomplished and your passion for your work. Your dedication to science and the pursuit of truth has inspired me to strive for greatness. This dissertation would not have been possible without your wisdom and encouragement, and for that, I offer my deepest gratitude.

I extend my deepest gratitude to my advisor, Dr. Marina Sirota, for granting me the opportunity to lead these two challenging yet transformative research projects. Your unyielding support was indispensable to the successful completion of my doctoral studies. I also wish to express my sincere thanks to my thesis committee members, Dr. Yadong Huang and Dr. Michael Keiser, for their invaluable scientific collaboration, mentorship, and constant support. Your insights and guidance have been critical to my academic and research achievements. Additionally, I am profoundly grateful to the members of the Sirota and Huang laboratories for their stimulating scientific discussions and supporting friendships, which significantly enriched my graduate experience.

CONTRIBUTIONS

For chapter 2:

Yaqiao Li, Carlota Pereda Serras, Jessica Blumenfeld, Min Xie, Yanxia Hao, Elise Deng, Lisa Chun, Julia Holtzman, Alice An, Seo Yeon Yoon, Antara Rao, Sarah Woldemariam, Alice Tang, Alex R. Zhang, Jeffrey Simms, Iris Lo, Tomiko Oskotsky, Michael J Keiser, Yadong Huang and Marina Sirota

For chapter 3:

Yaqiao Li, Yanxia Hao, Carlota Pereda Serras, Seo Yeon Yoon, Brian Grone, Leonardo Ding, Jessica Blumenfeld, Tomiko Oskotsky, Yadong Huang and Marina Sirota

A SYSTEMS BIOLOGY APPROACH TO PRECISION MEDICINE FOR ALZHEIMER’S DISEASE: CELL-TYPE-DIRECTED NETWORK-CORRECTING THERAPEUTICS AND TRANSCRIPTOMIC PROFILING ACROSS DISEASE RISK FACTORS

YAQIAO LI

ABSTRACT

Alzheimer's disease (AD) is a multifactorial neurodegenerative condition characterized by heterogeneous molecular alterations across various brain cell types, posing significant challenges for the development of effective treatments. To address this complexity, we undertook two complementary projects aimed at advancing precision medicine approaches for AD, each building on the other to create a robust foundation for targeted therapies.

The first project serves as a proof of concept for an innovative drug discovery strategy, utilizing a network correction approach grounded in direct human evidence and real-world data. By integrating diverse datasets, including single-cell human transcriptomics, drug perturbations, and electronic clinical records, we identified the combination of letrozole and irinotecan as potential therapeutics designed to correct gene expression alterations across multiple cell types implicated in AD. Rigorous validation in AD mouse models demonstrated that this combination therapy, targeting both neurons and glial cells, significantly ameliorated memory deficits and other AD-related pathologies, outperforming the single-drug treatments targeting either neurons or glial cells alone. The success of this project underscores the potential of cell-type-directed network-correcting therapy, demonstrating that targeting the transcriptomic landscape at a cell-type-specific level may offer a more efficacious approach to treating AD.

Building on this foundation, the second project was designed to expand the knowledge foundation for precision medicine by comprehensively characterizing the molecular influences of major AD risk factors, such as age, apolipoprotein E4 (*APOE4*), and sex using AD mouse models. By analyzing single-nucleus RNA-sequencing (snRNA-seq) data from the hippocampus of human *APOE4* and *APOE3* knock-in (KI) female and male mice across different ages, we identified significant variations in cell type abundance and gene expression patterns, particularly driven by sex differences and the interplay between age and *APOE* genotype. This detailed molecular profiling not only enriches our understanding of the disease but also provides a valuable dataset for future applications of the network correction method. Specifically, it enables the mapping of distinct risk profiles and the identification of tailored therapeutic interventions for individuals based on their unique transcriptomic signatures.

Together, these two projects are highly complementary: the first project validates the efficacy of network correction therapy, while the second project supplies the essential molecular data needed to understand disease heterogeneity, enabling the application of this therapeutic approach in a precision medicine framework. By integrating these insights, we can more accurately tailor treatments to individuals with distinct risk profiles, thereby advancing the development of personalized therapies for AD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Challenges in Drug Discovery for Alzheimer's Disease

Alzheimer's disease (AD) is an irreversible neurodegenerative disorder that clinically manifests in progressive memory decline, behavior changes, and cognitive and motor deficits[1]. Patients in advanced stages suffer from hallucinations, disorientation, and lack of self-sufficiency, and eventually die from severe respiratory failure, infection, or fasting[2]. AD is the most common form of dementia and reaching epidemic proportions, as it affects millions of patients globally[3]. The prevalence of AD is projected to triple by 2050[3]. Consequently, AD ranks among the leading causes of disability and dependency among older adults, placing significant strain on caregivers and healthcare systems[4]. Given its profound impact, urgent action is required to develop treatments that are both more effective and accessible.

Despite the growing prevalence and substantial economic burden, there is no cure currently exists for AD. There are only a handful of treatment options available to AD patients, and they have minimal impact on the severity and progression of the disease[5]. Pharmacological treatments include cholinesterase inhibitors, such as donepezil, rivastigmine, and galantamine, which help manage symptoms by increasing levels of acetylcholine in the brain, and memantine, which regulates glutamate to improve cognition and function[6]. These treatments primarily focus on diminishing disease symptoms and do not address the underlying causes. Recently, disease-modifying therapies like aducanumab and lecanemab, which target amyloid plaques, have been approved for certain patients[6]. However, aducamumab is discontinued in 2024 and lecanemab was showed to slows cognitive decline by about 25%[7]. Additionally, the extend of treatment efficacy varied greatly among patients with considerable side effects that are concerning[8]. Given these limited successes, developing more effective disease-modifying treatments for AD targeting specific pathophysiological mechanisms is urgently needed.

Over the past two decades, more than 2,700 clinical trials targeting individual core pathologies of AD were initiated, yet only two have received FDA approval[9]. This underscores the profound challenges in AD drug development, primarily due to the multifactorial and heterogeneous nature of the disease. The pathogenesis of AD is inherently multifactorial, characterized by a complex and dynamic interplay of pathogenic processes that occur across a diverse array of brain cell types[10,11]. The most prominent pathological features of AD are the accumulation of extracellular amyloid-β (Aβ) plaques and intracellular tau protein neurofibrillary tangles (NFTs)[12]. These aggregates are accompanied by impaired neurotransmitter signaling, synaptic dysfunction, dystrophic neurites, dysregulated neuronal metabolism, oxidative stress, cellular senescence, and cerebral atrophy[2,12]. Emerging evidence also highlights the critical role of neuroinflammation, and cellular dysfunction led by glial cells—microglia, astrocytes, and oligodendrocytes—in AD pathophysiology[13–15]. These glial cells, essential for maintaining brain homeostasis, not only lose their normal functions in AD but also exhibit pathological responses that produce reactive nitrogen species and reactive oxygen species, contributing to oxidative stress and subsequent neuronal death[13,15]. Given this complexity, effective treatment strategies for AD will likely need to address pathological changes across multiple disease-relevant cell types.

In Chapter 2, we detailed an initiative aimed at developing a comprehensive therapeutic approach using a combination of two drugs, each specifically targeting molecular alterations associated with AD in either neurons or glial cells. This strategy is designed to address the multifactorial complexity of AD by employing a multi-target, network-correcting treatment paradigm that accommodates the diverse and interrelated disease mechanisms occurring across multiple cell types implicated in the pathogenesis of AD. This approach underscores the importance of targeting the intricate cellular networks that drive the progression of the disease, offering a more holistic and potentially effective therapeutic solution.

AD is characterized by significant heterogeneity, encompassing various aspects, including clinical, genetic, and pathological variability, as well as differences in drug response, sex, and age of onset. Clinically, AD manifests through diverse symptomatology, with patients exhibiting different subsets of symptoms, such as memory loss, language difficulties, and behavioral changes[16]. The progression of cognitive decline also varies, with some individuals experiencing rapid deterioration while others show a more gradual decline[16]. Genetically, familial AD is linked to mutations in the amyloid precursor protein (APP) and presenilin genes, while sporadic AD is associated with risk variants in genes like apolipoprotein E (*APOE*) and triggering receptor expressed on myeloid cells-2 (TREM2), which are particularly enriched in glial cells[17]. Pathologically, AD is associated with Aβ plaques and tau tangles in the brain, yet the distribution and density of these abnormalities can differ widely[12]. Some individuals may exhibit high amyloid levels with minimal cognitive impairment, while others experience significant cognitive decline despite a lower amyloid burden[16]. Moreover, patient responses to treatment vary, with some benefiting from specific medications while others do not, reflecting underlying differences in disease mechanisms[18]. Age of onset further contributes to AD heterogeneity, with early-onset AD (before age 65) often having a stronger genetic component, and late-onset AD (after age 65) being more prevalent and influenced by a combination of genetic and environmental factors[19]. Additionally, AD exhibits prominent sex differences in prevalence, pathology, and symptomatology. About two-thirds of all AD patients are females[16]. Women tend to experience a faster rate of cognitive decline and score lower on verbal memory tests compared to men[20]. Although pathological biomarker discrepancies are observed between sexes, the same pathological burden often has more detrimental effects on cognition in women than in men[21]. The extensive heterogeneity of AD presents a formidable challenge, underscoring the importance of understanding these variations to develop personalized treatments and interventions, as a one-size-fits-all approach is unlikely to be effective given the diverse nature of the disease.

In Chapter 3, we sought to address the inherent heterogeneity of AD, thereby paving the way for the future application of precision medicine in its treatment. Specifically, we conducted a systematic investigation into the molecular influences of three predominant AD risk factors: age, sex, and *APOE* genotypes. Through this comprehensive analysis, we elucidated the distinct molecular landscapes shaped by these factors, thereby establishing a foundation for the development of tailored therapeutic interventions. This approach enables the adaptation of network correction strategies to align with the unique transcriptomic profiles of individual patients, ultimately advancing the potential for personalized and more effective treatment modalities in AD.

1.2 Characterizing Alzheimer's Disease Through the Interplay of Major Risk Factors

Age, apolipoprotein E ε4 (*APOE4*), and chromosomal sex are well-established non-modifiable risk factors for late-onset AD[22]. These inherent characteristics cannot be altered or influenced by individuals, yet they play a crucial role in shaping an individual's risk profile. Leveraging non-modifiable risk factors offers significant advantages in the development of personalized medicine. Unlike lifestyle and environmental factors, genetic risk variants can be directly and reliably measured through genetic testing, enabling the early detection of those at higher risk and facilitating targeted interventions before the onset of symptoms. Additionally, the advanced and personalized biological characterization of complex diseases allows for the classification of individuals into subgroups with distinct risk profiles, which in turn enables tailored therapeutic approaches based on their unique genetic and biological makeup. This stratification is not only critical for optimizing treatment effectiveness but also for reducing potential side effects by selecting the most appropriate therapies for each patient. Understanding the influence of these non-modifiable risk factors on disease pathophysiology facilitate the development of targeted therapy and integrating this information into patient care is essential for healthcare providers, as it allows for the delivery of more precise, effective, and individualized prevention and treatment strategies. This comprehensive approach ultimately leads to improved health outcomes, especially in managing complex, multifactorial diseases like AD, where early and tailored interventions can make a significant difference.

Aging is the most significant risk factor for AD and serves as a fundamental driver in the development of the disease[23]. The incidence and prevalence of AD increase markedly with advancing age[24]. Approximately 80% of AD patients are over the age of 75, with the incidence rising from 2 per 1,000 individuals between the ages of 65 and 74 to 37 per 1,000 individuals aged 85 and older[25]. Moreover, the aging of the global population is occurring at an unprecedented rate, with the number of older adults worldwide projected to more than double by 2050[3]. This demographic shift underscores the urgent need for more effective and accessible treatments to address the growing burden of AD.

The age of onset of AD can be classified into early-onset (before age 65) and late-onset (after age 65)[19]. Early-onset AD often has a stronger genetic basis, including mutations in the APP, PSEN1, and PSEN2 genes, and tends to be more aggressive[26]. Late-onset AD, which is more common, is influenced by a combination of genetic, environmental, and lifestyle factors[24]. Age-related changes in the brain, such as the accumulation of amyloid plaques and tau tangles, as well as reduced neuroplasticity and increased oxidative stress, contribute to the disease[12]. The interaction between aging processes and genetic risk factors, such as *APOE*, further complicates the disease trajectory.

Apolipoprotein E (APOE) is a critical protein involved in lipid metabolism, primarily responsible for the transport and redistribution of lipids, such as cholesterol and triglycerides, among various tissues in the body[27]. It plays several essential roles in maintaining lipid homeostasis, particularly in the central nervous system (CNS) and cardiovascular system[27]. The *APOE* gene, particularly the ε4 allele, is the strongest genetic risk factor for late-onset AD[17]. Over 60% of individuals with AD carrying at least one *APOE4* allele[22]. Individuals with one copy of the ε4 allele have a 2-3 times higher risk of developing AD, while those with two copies have a 10-15 times higher risk[28]. In contrast, the *APOE* ε2 allele is thought to have a protective effect against the disease[28]. APOE influences the aggregation and clearance of amyloid-beta, a key pathological hallmark of AD[11]. The APOE4 isoform is less efficient at promoting amyloid-beta clearance compared to APOE2 and APOE3, leading to a greater accumulation of amyloid plaques in the brain[11]. In addition to amyloid-beta, APOE, particularly the APOE4 isoform, is implicated in tau pathology[29]. *APOE4* carriers are more prone to tau aggregation and neurofibrillary tangle formation, which contribute to neurodegeneration and cognitive decline in AD[30]. APOE also plays a role in modulating the immune response and inflammation, particularly within the CNS. It influences the activity of microglia, the brain's resident immune cells, affecting their ability to respond to injury and clear amyloid-beta[11]. The APOE4 isoform is associated with a more pro-inflammatory state, which may contribute to the chronic neuroinflammation observed in AD[31].

Sex differences have been clinically documented in AD[23], yet the underlying cause for these differences are not well understood. Women are disproportionately affected by AD and progress at faster rate of cognitive decline than men[21]. This accelerated decline is observed particularly in memory-related functions, where women typically outperform men in early stages but decline more rapidly as the disease progresses[16]. In addition to greater longevity in females[32], other biological differences may be responsible for the higher prevalence and accelerated cognitive decline observed in women during disease progression[33]. For instance, a longitudinal study examining a postmortem cohort of about 1,500 individuals observed that in the presence of similarly high Aβ burden, females exhibited faster cognitive decline than males[34], suggesting females might be more susceptible to Aβ toxicities. Furthermore, after adjusting for age and education, women had a higher tau tangle density[35].

While studying individual risk factors allows for precise, controlled investigations that contribute to a foundational understanding of disease mechanisms, this approach often fails to capture the complexities of how these factors interact in real-world scenarios. Investigating multiple risk factors in tandem provides a more comprehensive view of disease etiology, reflecting the dynamic interplay that occurs in actual biological contexts. Numerous studies have emphasized the significant interplay between age, sex, and *APOE* genotype in shaping the risk, progression, and clinical manifestation of AD[36]. For instance, the association of *APOE4* on AD might be age dependent. *APOE4* exerts its maximal effect on AD-risk by early 70’s with a reduction in risk after age 85 in both sexes. The impact of the *APOE* ε4 allele on AD risk is more pronounced in women than in men. Women who carry the ε4 allele are more likely to develop AD and tend to show a faster rate of cognitive decline compared to male carriers. The reasons for this difference are not fully understood but may involve hormonal influences, such as the loss of estrogen during menopause, which interacts with APOE and affects brain function. While substantial effort has been dedicated to investigating this interplay, deciphering these intricate interactions remains highly challenging, especially when analyzing the convergence of multiple risk factors across diverse data domains. The increased complexity in study design, data collection, and interpretation complicates the task of fully understanding these relationships. Despite these challenges, a systems biology approach, which examines the interaction between *APOE* genotypes and sex across different stages of aging, offers a promising avenue for studying disease-associated transcriptomic alternations with each risk factor in a cell-type-specific manner. By leveraging advanced bioinformatics tools and large-scale data analytics, researchers can integrate and analyze complex datasets, providing a robust framework for understanding these multifaceted interactions transcriptomically. This approach ultimately holds the potential to lead to more effective and targeted interventions for AD by correcting the distinct pathological transcriptomic landscape in each specific patient population.

1.3 A System Biology Approach to Precision Medicine

Single-cell technology represents a transformative advancement in biological research, offering unprecedented insights into cellular diversity, function, and behavior at an individual cell level. Unlike traditional bulk analyses, which average signals across large populations of cells, single-cell technology allows researchers to dissect the heterogeneity within tissues, revealing the intricate details of cellular processes and interactions.

The generation and availability of single-cell RNA-sequencing (scRNA-seq) datasets from AD brain samples and mouse models of AD[31,37–39] offer novel opportunities to unravel the cellular heterogeneity underlying disease mechanisms. This technological advancement enables the application of cellular precision computational drug screening for AD, leveraging transcriptomic databases of drug perturbations[40]. Single-cell transcriptomes facilitate gene expression profiling at the individual cell level, capturing the dynamic molecular properties and actions distinguishing diseased cells from their healthy counterparts. Beyond assessing the molecular perturbations across thousands of genes, scRNA-seq also enables the identification of specific cell types and the origins of cellular heterogeneity, making it particularly suited for studying AD's complex cellular mechanisms[31].

In conjunction with disease-specific transcriptomic profiles, a computational drug repositioning pipeline, previously established in the Sirota lab, utilizes drug-induced expression profiles of thousands of existing drugs and small molecule compounds provided by the Connectivity Map (CMap) to uncover associations among genes, diseases, and potential drug candidates[41]. The drug repositioning algorithm is predicated on the hypothesis that drugs capable of reversing gene expression patterns in a manner opposite to that observed in the disease state may exhibit therapeutic efficacy. Essentially, the selected drug candidates are assumed to modulate the transcriptomic landscape of the diseased state, aligning it more closely with the gene expression profile observed in healthy control individuals. Another essential feature of the rationale behind the computational drug screen pipeline is the ability to leverage the multi-target effect of drugs to identify drug that has a network perturbing effect. A single drug might interact with multiple molecular targets within a tissue system, and lead to a network of gene expression changes. By mapping the pharmacological effects and mechanisms of action within a drug network, it is possible to rationally interpret and identify top-scoring drugs predicted to be effective for each cell type involved in the disease. This approach holds promise for developing effective disease-modifying treatments by combining drugs that target the heterogeneous cellular landscape of AD.

The CMap represents the most comprehensive transcriptomic drug profile library currently available, predominantly encompassing FDA-approved or previously investigated compounds[42]. While CMap provides an extensive resource, its drug profiles were primarily generated in cancer cell lines, which may not accurately represent the transcriptomic landscape of brain tissues. Ideally, drug profiles derived from a wide array of compounds tested on different brain cell types would be more appropriate for AD research; however, such data are currently lacking. Given that most compounds in the CMap are existing drugs, the top predicted candidates are ideal for repositioning toward AD.

Drug repositioning has gained interest in AD therapeutic research due to its accelerated process in drug development[43] and the fact that it enabled the identification of successful therapies for many diseases including cancer[44] and Parkinson's disease (PD)[45]. Several repurposing approaches have been explored for AD, including candidate selection based on drug-associated patient outcomes, known on or off-targets of existing drugs, as well as matching drug mechanisms to disease-associated transcriptomic signatures[46]. However, multiple attempts at drug repositioning for AD treatment targets single disease-specific molecular pathways or pathological hallmarks and leaving the multifactorial nature of AD unaddressed[43]. As we learned that AD likely involves cellular heterogeneous pathological features, computationally repurposed combination therapies targeting a network of disease-modifying molecular mechanisms implicated in different brain cell types might be more effective against AD.

1.4 Leveraging Real-World Data to Repurpose Existing Drugs as Novel Alzheimer’s Disease Therapeutics

The integration of real-world data into the drug discovery process offers the potential to significantly improve the success rate of developing effective therapies. Unlike traditional clinical trial data, which is often limited in scope and obtained under controlled conditions, real-world data is derived from a variety of sources, including electronic medical records (EMRs), and reflects the actual experiences of diverse patient populations in everyday clinical settings[9]. This extensive dataset provides a more comprehensive understanding of how drugs perform across different populations, particularly those typically underrepresented in clinical trials. By capturing a broader range of patient demographics, comorbidities, and treatment patterns, real-world data can reveal new insights into drug efficacy, safety, and potential off-target effects, thereby enhancing decision-making throughout the drug development pipeline. Moreover, real-world data allows for the identification of associations between drug usage and the disease of interest, facilitating the discovery of novel therapeutic targets and the repurposing of existing drugs. These insights can also support the design of more personalized treatment strategies, ultimately leading to more effective and tailored therapies.

EMR offer a rich and valuable source of real-world data that can significantly aid in identifying candidate drugs for AD[47]. By analyzing the comprehensive clinical data within EMRs, researchers can uncover patterns of drug use and health outcomes in large populations, facilitating the identification of drugs that may have off-target effects beneficial for AD[48]. For instance, EMR data can reveal associations between certain medications and a lower incidence of AD or slower disease progression in patients who were originally prescribed these drugs for other conditions. This observational approach, known as drug repurposing, allows researchers to identify potential AD therapeutics from medications already approved for other indications, thereby reducing the time and cost associated with drug development. Additionally, EMRs provide longitudinal data, enabling the analysis of long-term effects of drug exposure on cognitive health, including the onset and progression of AD symptoms. By integrating EMR data with sex and genomic information, such as *APOE* genotype, researchers can further stratify patients to identify subgroups that respond differently to certain drugs, paving the way for more personalized therapeutic strategies. Moreover, advanced data mining techniques, such as machine learning algorithms, can be applied to EMR datasets to identify complex interactions between drugs, comorbidities, and patient demographics that might not be apparent through traditional analysis. Overall, the systematic analysis of EMR data offers a powerful tool for discovering and validating candidate drugs for AD, leveraging real-world evidence to drive more effective and targeted interventions.

Mapping the transcriptomic landscape using large-scale single-cell datasets, combined with understanding the heterogeneity of AD through molecular and clinical data, represents a promising approach to establishing precision medicine for AD. Single-cell transcriptomics allows for the detailed characterization of gene expression profiles at the individual cell level, revealing the diversity of cellular states and the complex interactions among various cell types implicated in AD. By integrating this granular molecular data with comprehensive clinical information—such as patient demographics, disease progression, and genetic risk factors—researchers can identify specific subpopulations of patients who share similar molecular and clinical profiles. This approach enables the stratification of patients based on their unique disease mechanisms, facilitating the design of tailored therapeutic strategies that target the distinct biological pathways driving disease in each subgroup. Ultimately, this integration of single-cell transcriptomics and clinical heterogeneity offers a powerful framework for advancing precision medicine, providing more personalized and effective treatments for patients with AD.

CHAPTER 2: DATA-DRIVEN DISCOVERY OF CELL-TYPE-DIRECTED NETWORK-CORRECTING COMBINATION THERAPY FOR ALZHEIMER'S DISEASE

2.1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder with severe impacts on individuals, families, and society[1], and yet without a cure. AD patients suffer progressive memory loss, behavior changes, and cognitive and motor deficits, leading to diminished quality of life[50]. Nearly 30 million elderly people currently live with AD worldwide, a number projected to triple by 2050[51,52]. With global costs exceeding $1 trillion annually, AD is one of the costliest health conditions worldwide[53]. Urgent action is needed to develop effective and accessible treatments.

Despite rigorous preclinical and clinical research efforts, AD drug development faces significant challenges, with a 98% of 413 AD clinical trials conducted in the last decade failed to produce a drug[9]. Current treatments are mostly limited to symptom-managing treatments[6]. Furthermore, recently approved immunotherapies have only modest effects on disease progression[7]. The lack of effective treatments stems from the pathological heterogeneity in AD. While AD's most prominent disease hallmark is proteopathy, characterized by extracellular amyloid-β (Aβ) plaques and intracellular tau neurofibrillary tangles (NFTs), their interplay and exact mechanisms leading to disease remain unclear[12]. Genetic heterogeneity further complicates the disease, including risk mutations in the amyloid precursor protein (*APP*) and presenilin genes, and APOE4, a risk isoform of apolipoprotein E (*APOE*)[17]. Emerging evidence highlights the critical roles of different brain cell types, with neuroinflammation and inadequate neuronal support from malfunctioning glial cells contributing to AD progression[10]. Considering the multifactorial nature of AD, traditional therapeutic approaches focusing on single disease hallmarks or bulk-tissue-level pathologies are often insufficient, leading to variable treatment outcomes.

Given the unmet need for disease-modifying treatments, drug repurposing has gained interest due to its faster development, lower costs, and improved safety[46]. In addition, technical advancements in mining large-scale databases offer new opportunities for discovering promising candidates. These developments, combined with various drug screening approaches such as *in vitro*[54] and *in vivo* assays[55], gene signature matching[40], network modeling[56], machine learning[57], and data mining[58], have identified numerous repurposed candidates for AD over the past decade. However, the extensive array of potential candidates complicates the establishment of priorities for clinical translation[43].

In this chapter, we propose repurposing a combination of two existing drugs, letrozole[59] and irinotecan[60], to reverse cell-type-specific gene expression alterations across multiple cell types implicated in AD. Our drug screening strategy is entirely driven by human data, utilizing large-scale omic datasets from post-mortem brains[37–39], a drug perturbation library generated in human cell lines[42], and clinical records encompassing millions of individuals, thereby maximizing the chance of clinical translation. We validated our predicted drug candidates through dosing experiments in an AD mouse model[61,62], showing that the combination therapy targeting both neurons and glial cells significantly ameliorated memory deficits and AD-related pathologies compared to vehicle treatment, and outperformed single-drug treatments targeting either neurons (letrozole) or glial cells (irinotecan) alone. The successful *in vivo* validation underscores the power of multi-cell-type network-correction therapies in effectively treating AD, highlighting the promise of our human data-driven, cell-specific drug discovery approach to develop innovative therapies for complex diseases. Furthermore, by targeting molecular signatures and clinical features derived from real-world individuals, this study demonstrates the potential for AI-enabled precision medicine leveraging large-scale multimodal personalized measurements.

2.2 Results

We organized a comprehensive single nucleus RNA-sequencing (snRNA-seq) dataset by combining published data from three independent studies[37–39](**Figure 2.4.1**), covering 37 AD patients (15 females and 22 males) and 29 controls (13 females and 22 males). Samples from individuals who did not meet both CERAD[63] and Braak[64] criteria for AD or control were excluded from our dataset. The Uniform Manifold Approximation and Projection (UMAP) visualization reveals clustering predominantly by the datasets of origin rather than biological variations, such as cell types or disease status (**Figure 2.4.1, Figure 2.4.7**). After batch harmonization with the integration algorithm, technical artifacts between datasets were effectively removed (**Figure 2.4.1**), and distinct cell populations clustered by cell types (**Figure 2.4.1**) while no discrete separation was observed based on disease status (**Figure 2.4.1**). The integrated dataset consisted of expressions of 29,120 features in 137,065 cells.

To systematically characterize cell-type-specific AD pathophysiological features, we conducted comprehensive analyses focusing on six disease-relevant cell types: excitatory neurons (ex.neu), inhibitory neurons (in.neu) microglia (mic), astrocytes (ast), oligodendrocytes (oli), and oligodendrocyte precursor cells (OPC). In AD patients, the proportion of inhibitory neurons significantly decreased compared to controls (**Figure 2.4.1**), as previously reported in human and mouse models of AD[27,31,65]. Cell-cell-communication analysis revealed heterogeneous signaling patterns among neuronal subpopulations in AD compared to controls. Signaling to inhibitory neurons increased from all cell types except microglia, while signaling to excitatory neurons decreased from all cell types except oligodendrocytes (**Figure 2.4.1**).

Cell-type-specific differential gene expression analysis between AD and control groups revealed significant variations in differentially expressed genes (DEGs) across cell types (**Figure 2.4.1**). Many DEGs were unique to specific cell types, while others were shared but displayed opposite regulatory patterns in AD (**Figure 2.4.1**). For example, the *APOE* gene was significantly upregulated in microglia but downregulated in astrocytes and OPCs in AD samples. Additionally, gene set enrichment analysis showed that AD-enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and gene ontology (GO) terms exhibited extensive cell-type heterogeneity, with most pathways and terms unique to individual cell types (**Figure 2.4.1, Figure 2.4.7**). While pathways like chemical carcinogenesis - reactive oxygen species and prion disease were enriched in AD across all six cell types, unique pathways delineated them into two clusters. The neuronal-centric cluster included excitatory and inhibitory neurons, with oligodendrocytes sharing estrogen signaling with excitatory neurons and cAMP signaling with inhibitory neurons. The glial-centric cluster was comprised of astrocytes, microglia, and OPCs, with astrocytes and OPCs sharing gap junction pathway enrichment, and microglia and OPCs enriched for oxidative phosphorylation. These findings suggest AD pathogenesis involves heterogeneous transcriptomic-driven molecular alterations manifested in discordant behaviors of multiple cell types. Effective treatment likely needs to correct malfunctions in multiple, if not all, cell types.

After establishing cell-type-specific AD profiles from our integrated human dataset, we screened for network-correcting drug candidates targeting AD-specific transcriptomic changes across multiple cell types. We queried each cell-type-specific AD transcriptomic profile against the Connectivity Map (CMap) drug expression database[42] using a computational pipeline that matches gene expression profiles of diseases and existing drugs (**Figure 2.4.2**). Most AD signatures overlapped with drug profile features, which were then used as inputs in the pattern-matching algorithm (**Figure 2.4.2**). With a false discovery rate (FDR) < 0.05, we predicted 35 hits for excitatory neurons, 12 hits for inhibitory neurons, 8 hits for microglia, 33 hits for astrocytes, 4 hits for oligodendrocytes, and 8 hits for OPCs. Several drug hits overlapped across cell types, forming linkages in a network visualization of drug interactions with cell types (**Figure 2.4.2**).

Notably, 25 repurposed drugs significantly reversed cell-type-specific AD profiles in multiple cell types (**Figure 2.4.2**), indicating a multi-targeted potential with these drugs. These multi-cell-type drug candidates span various therapeutic classes, including cardiac glycosides (digitoxigenin, digoxigenin, helveticoside), chemotherapeutic agents (methotrexate, irinotecan, etoposide, mitoxantrone), aromatase inhibitors (letrozole), histone deacetylase inhibitors (trichostatinA, vorinostat, scriptaid), immunosuppressants/mTOR inhibitors (sirolimus), anti-inflammatory/anti-tumor agents (15-delta prostaglandin J2), nonsteroidal anti-inflammatory drugs (fenbufen), antifungal agents (ciclopirox), antibiotic (monensin, ionomycin), antiepileptics (valproic acid), antipsychotics (haloperidol), and several experimental compounds with undefined pharmacological classifications (LY-294002, CP-320650-01, CP-690334-01, syrosingopine, hycanthone, 5155877)[66].

We aimed to validate the effects of the repurposed drug candidates in humans using real-world data. An advantage of existing pharmaceutical agents is their potential for population-based analyses using real-world patient data to investigate associations between drug use and AD outcomes. We identified 25,257 individuals diagnosed with AD in the University of California (UC)-wide Electronic Medical Records (EMR) database among 1.4 million individuals aged 65 or older, constituting a substantial clinical database for analysis.

Focusing on the 25 multi-cell-type candidate drugs, we found usage records for 10 of these drugs, as only a subset of the drugs from CMap is FDA-approved or prescribed (**Figure 2.4.2, Table 2.5.3**). Of these ten drugs, five (letrozole, irinotecan, methotrexate, ciclopirox, and sirolimus) were associated with a significantly reduced risk of AD compared to matched controls, suggesting potential protective effects of them against AD. Although three drugs (etoposide, trifluoperazine, and vorinostat) showed reduced risk, statistical significance was not achieved due to insufficient patient representations. Lastly, two drugs (valproic acid and haloperidol), both used for neurological conditions, showed higher relative risk scores.

We hypothesized that combination therapy targeting both neuronal and glial transcriptomic profiles might more effectively alleviate AD pathologies. Among five drug candidates showing significant AD risk reduction in UC-wide EMR, we prioritized letrozole for its predicted reversal effects in excitatory and inhibitory neurons, and irinotecan for its effects on the glial-centric cluster, including astrocytes, microglia, and OPCs (**Figure 2.4.2**). Therefore, a combination of letrozole and irinotecan potentially targets five cell types in AD.

When visualizing AD and drug profiles side by side, genes upregulated in AD neurons shifted downward, while downregulated genes shifted upward in letrozole-treated profiles. Irinotecan-treated profiles showed similar shifts across glial cells (**Figure 2.4.2**). Letrozole, an aromatase inhibitor primarily prescribed for breast cancer treatment and occasionally for male infertility[59,67], had a relative risk ratio of 0.466 for AD in UC-wide EMR. It demonstrated reversal scores of -0.37 and -0.61 for excitatory and inhibitory neurons, respectively. Despite a female-skewed cohort (33 females to 1 male), the risk reduction ratios were comparable for both sexes when analyzed individually (**Table 2.5.4**). Irinotecan, a DNA topoisomerase I inhibitor used for colorectal cancer treatment[60], showed a relative risk ratio of 0.195 for AD. It demonstrated AD reversal capabilities across all three cell types of the glial-centric clusters, with reversal scores of -0.29, -0.31, and -0.45 for astrocytes, microglia, and OPCs, respectively.

To experimentally test the efficacy of the combination therapy with letrozole and irinotecan, we conducted dosing experiments in an AD mouse model expressing mutant human APP/PS1 and tau (5xFAD[62] x PS19[61]), which recapitulates many AD-related phenotypes including amyloid plaque formation, tau tangles, and gliosis, with an early and aggressive pathology onset[68,69] (**Figure 2.4.3**). A sex-balanced double transgenic cohort was evenly divided into four groups (n=20, including both sexes) and each treated with either vehicle, letrozole, irinotecan, or the combination of both drugs every other day for three months. Spatial learning and memory performance were evaluated using the Morris Water Maze (MWM) test[70]. While no significant difference was observed in learning trials across 6 days between each treatment and vehicle group (**Figure 2.4.3**), the memory test in probe trials revealed that only the combination-treatment group exhibited a statistically significant preference for the target quadrant at both 24 and 72 hours post-learning trials (**Figure 2.4.3**), suggesting rescue of both short-term and long-term memory deficits by the combination treatment. Furthermore, only the combination-treatment group demonstrated significantly better location recall by making more frequent crossings of the platform location at both time points after platform removal (**Figure 2.4.3**). Mice in the combination-treatment group also significantly outperformed the vehicle mice in latency to the first platform crossing during the 72-hour probe trial (**Figure 2.4.8**). Average swim speeds were similar across all groups, indicating that the observed behavioral differences were not confounded by physical performance or motor function (**Figure 2.4.8**).

Additionally, sex differences were observed in the dosing experiment, with significantly improved learning performance only observed in combination-treated males compared to vehicle-treated males, and not in females (**Figure 2.4.8**). Memory rescue by the single drug treatments was also evident in males at 24 and 72 hours of the probe trials but not in females (**Figure 2.4.8**). Taken together, AD-related behavior assessments validated the efficacy of the combination therapy with letrozole and irinotecan, especially in males, in an AD mouse model with both amyloid and tau pathologies, and single-drug treatments alone had much less efficiency.

To determine whether the single or combination drug treatments can also rescue AD-related pathologies, we morphologically assessed the cohort at 9 months of age (4 months post-treatment). According to previous literature, this AD mouse model develops extensive neurodegeneration, Aβ deposits, hyperphosphorylated tau (p-tau), gliosis, and severe loss of CA1 neurons in the hippocampus at this age[68,69]. We first evaluated neurodegeneration in the hippocampal region, as atrophy in this region is a hallmark of AD progression. Analyses of hippocampal volume showed rescue of atrophy in all treatment groups, with the most significant improvement in the combination-treatment group compared to the control group (**Figure 2.4.4**). Thioflavin S (Thio-S) staining for β-amyloid pathology revealed significant reductions in all treatment groups compared to vehicle-treated controls, measured by the Thio-S-positive percent area and plaque counts normalized by hippocampal size (**Figure 2.4.4**). P-tau pathology was assessed by immunofluorescent staining using the p-tau-specific AT8 antibody. While all treatment groups showed a trend of reduced p-tau pathology compared to vehicle-treated controls, only the combination-treatment group had a statistically significant reduction in the p-tau coverage area of the hippocampus (**Figure 2.4.4**).

Given the cell-type-precision therapeutic design, we investigated the effects of each treatment on gliosis and neuronal loss through immunostaining of specific cell types. As major contributors to neuroinflammation, microgliosis and astrogliosis were evaluated by quantifying the coverage area of cell type-specific markers Iba1 and GFAP, respectively. Mice in irinotecan and combination-treatment groups showed a significant reduction in the Iba1-positive percent area in the hippocampus, indicating alleviation of microgliosis (**Figure 2.4.4**). Astrogliosis reduction was moderate, with significant reduction observed only in the irinotecan-treated group (**Figure 2.4.4**).

Neuronal loss was assessed by measuring neuronal layer thickness via NeuN staining in the CA1 region of the hippocampus. Significant rescue of neuronal loss was evident in the CA1 region (**Figure 2.4.4**). The most significant rescues were observed in letrozole and combination-treatment groups. Interestingly, irinotecan-treated mice also showed moderate rescue in neuronal loss, likely attributable to the alleviation of gliosis that may help prevent neuronal loss.

To investigate the cell-type-specific transcriptomic changes in response to the combination treatment, snRNA-seq was performed on dissected hippocampi obtained from mice on combination treatment of letrozole and irinotecan (L+I) or vehicle treatment (n=8 per group, including both sexes). After standard processing and quality control, we obtained a filtered dataset containing 25,642 gene features expressed across 237,853 nuclei for further analysis. Through graph-based clustering and visualization using UMAP, we identified 31 distinct cell clusters, including clusters assigned to the six major cell types (**Figure 2.4.5**). Additionally, we found that the L+I treatment reduced the excessive communication between inhibitory neurons and other cell types (**Figure 2.4.5**), evident in AD (see **Figure 2.4.1**).

To account for the granularity and selective vulnerabilities among neuronal subtypes, we subdivided the excitatory and inhibitory neurons into CA1 pyramidal cells, CA3 pyramidal cells, dentate granule cells (DGC), subiculum neurons, and interneurons based on expressions of hippocampal subregion marker genes. Notably, in the UMAP plot split by treatment groups, there is a discernible higher density of cells observed in the L+I-treated group compared to the vehicle-treated group within the pyramidal neuron clusters of the CA1 and CA3 regions (**Figure 2.4.5**). This was quantified via cell type abundance analysis, revealing that the proportions of CA1 and CA3 pyramidal neurons in L+I-treated mice were significantly higher than in vehicle-treated mice (**Figure 2.4.5**), consistent with our findings from pathological analysis.

Differential gene expression analysis comparing L+I and vehicle treatments revealed varying counts and compositions of DEGs (abs(logFC) > 0.1, padj < 0.05) across cell types, with the fewest DEGs observed in oligodendrocytes, suggesting they were least affected (**Figure 2.4.5**). Gene set enrichment analysis of cell-type-specific DEGs revealed enrichments in disease-relevant functional pathways (**Figure 2.4.5**). Notably, the treatment enriched pathways related to nervous system development and synaptic activities in neurons. For example, in excitatory neurons, the dendrite morphogenesis pathway was enriched, while the axonogenesis pathway was enriched in inhibitory neurons. In glial cells, oligodendrocyte differentiation pathway was enriched in OPCs, and actin filament-based movement and synapse assembly pathways in oligodendrocytes. Essential microglial functions like pathways of response to oxidative stress and histamine, regulation of neuron death and synaptic plasticity, and neuron projection development were enriched, as were neuron development and regulation of synaptic activity pathways in astrocytes. KEGG pathway analysis revealed that L+I-treatment perturbed pathways associated with AD, such as long-term potentiation, circadian entrainment, cAMP signaling, and calcium signaling (**Figure 2.4.5**), suggesting that the combination treatment promotes neuroprotective functional pathways in a cell-type-specific manner.

Lastly, we explored drug treatment-reversed transcriptomic signatures of AD to elucidate potential molecular and cellular mechanisms underlying the observed benefits of the L+I treatment. We mapped human AD signatures to homologous mouse genes. With an absolute log fold change cutoff of 0.01, the L+I treatment reversed the expression patterns of AD signatures in multiple cell types: 56 (27%) in excitatory neurons, 36 (29%) in inhibitory neurons, 90 (27%) in astrocytes, 52 (20%) in microglia, 37 (15%) in oligodendrocytes, and 64 (34%) in OPCs (**Figure 2.6**). With a p-adj-value < 0.05 cutoff, 49 and 27 AD signatures were significantly reversed by L+I treatment in excitatory and inhibitory neurons, respectively, with 13 shared between them (**Figure 2.6**). In glial cells, fewer reversed gene expressions reached statistical significance: 13 in astrocytes, 6 in microglia, 11 in oligodendrocytes, and 19 in OPCs. However, they unveiled intriguing patterns. For instance, *APOE,* identified as a DEG in the human integrated analysis, was upregulated in microglia and downregulated in astrocytes and OPCs. The expression patterns of *Apoe* in the tested AD mouse model were reversed by L+I treatment across all three cell types (**Figure 2.4.6**), although statistical significance was only achieved in astrocytes and OPCs, possibly due to the presence of a small but diverse microglia population in the sequenced cohort.

We then conducted gene set enrichment analysis on significantly reversed genes to identify potential mechanistic targets. Combining reversed genes in excitatory and inhibitory neurons revealed associations with the estrogen signaling pathway among others (**Figure 2.4.6**). Reversed genes involved in this pathway also regulate protein kinase and phosphatase activity, along with tau kinase activity specifically, which could be directly relevant to AD. Furthermore, certain reversed genes are associated with synaptic activities and neuron projections, potentially contributing to the rescue of neurodegeneration noted in the pathological analysis. In astrocytes, reversed genes are associated with the regulation of long-term synaptic potentiation, chemical synaptic transmission, and cholesterol metabolism (**Figure 2.4.6**). In microglia, reversed genes are associated with pathways regulating synapses and cell growth (**Figure 2.4.6**). Additionally, we combined reversed genes from oligodendrocytes and OPCs, revealing associations with pathways related to cell growth regulation, response to reactive oxygen species, and neuron projection regulation (**Figure 2.4.6**). These findings provide a transcriptomic foundation, supporting the network correction concept that

L+I combination treatment rescues AD-related behavioral and pathological deficits by rectifying complex dysregulated gene networks across multiple disease-relevant cell types.

2.3 Methods

Human Single Nuclei RNA-Sequencing (snRNA-Seq) Data Curation

To ensure a diverse representation of AD patients, we curated publicly available human snRNA-seq datasets from three independent sources. Both Mathys et al. and Zhou et al. studies were obtained from the Accelerating Medicines Partnership Alzheimer's Disease Project (AMP-AD) Knowledge Portal (https://adknowledgeportal.org) under the Religious Order Study and Memory and Aging Project (ROSMAP). The Mathys et al. dataset is accessible through https://doi.org/10.7303/syn2580853. The Zhou et al. dataset is available under the study snRNAseqAD_TREM2 and is also accessible through https://doi.org/10.7303/syn21125841. Only individuals without TREM2 mutations were included in our integrated dataset. The third dataset by Lau et al. was obtained from Gene Expression Omnibus (GEO) under the accession number GSE157827.

Case-Control Standardization across Datasets

To standardize AD identification across studies, samples were re-classified into AD or control groups based on tau tangles severity with Braak clinical staging[64] scores and Aβ burden using Consortium to Establish a Registry for Alzheimer's Disease (CERAD) scores[63] as a proxy. Braak staging ranges from I to VI to represent low to high levels of tau deposition. The CERAD scoring ranges from 1 to 4 to indicate high to low Aβ burden severity. Based on the available metadata from the original studies, we defined AD cases as individuals with severe tau deposition (Braak ≥ IV) and high Aβ load (CERAD ≤ 2), and non-AD controls as individuals with low tau (Braak ≤ III) and low Aβ load (CERAD ≥ 3). Individuals with scores that did not fulfill both criteria were excluded. Our final integrated dataset consisted of 37 AD cases and 29 controls.

Human snRNA-Seq Dataset Integration, Normalization, and Batch Correction

Before merging the datasets, cells deemed as poor quality were removed with the following criteria: total feature count less than 500, total features less than 250, or mitochondrial gene ratio higher than 10%. Sparse features, expressed in fewer than 10 cells, were also removed. Simply merging datasets based on principle component analysis (PCA) reveals apparent study-associated variations. To perform batch harmonization while maximizing the conservation of disease-relevant biological variance, we performed canonical correlation analysis (CCA)[71] using the Seurat package v4.0.4 to harmonize the merged dataset. We applied the default settings to first log-normalize the merged dataset with Seurat's *NormalizeData* function. Then, for each dataset independently, the top 2000 variable features were identified using the *FindVariableFeatures* function with "vst" as the selection method. Integration anchors were identified with the *FindIntegrationAnchors* function based on the top variable features. Finally, a harmonized dataset consisting of 137,065 cells and 29,120 features was generated using the *IntegrateData* function in Seurat.

The data matrix was linearly transformed using Seurat's *ScaleData* function. Dimensionality reduction through Uniform Manifold Approximation and Projection (UMAP) was performed with the *RunUMAP* function, which considers the top 30 dimensions selected from the corresponding principal component analysis (PCA) obtained by running the *RunPCA* function in Seurat. Clustering was determined based on the first 30 PCs using the *FindNeighbors* function in Seurat, which embeds cells in a K-nearest neighbor graph based on Euclidean distance in PCA space and refines the edge weights between any two cells based on the shared overlap in their local neighborhoods. Clustering was implemented using a resolution of 0.5 in the *FindClusters* function, which applies modularity optimization techniques such as the Louvain algorithm, resulting in a set of 21 distinct clusters.

Human Cell Type Annotation

Cell type identities were determined by applying Seurat's AddModuleScore function to lists of known human brain marker genes (~ 8 per cell type) collected from PanglaoDB and referenced by Jiang et al. Cell type assignment included astrocytes, microglia, oligodendrocytes, oligodendrocyte precursor cells, endothelial cells, pericytes, and excitatory and inhibitory neurons. Each cell was assigned the corresponding cell type identity that generated the highest scores among scores for all cell types. If the highest and second highest scores of a cell were within 20% of the highest score, then the cells were deemed hybrids and excluded from further analysis. We assessed the validity of the assigned cell type identities by examining the homogeneity, distribution, and separation of cell types by clusterings in UMAP plots.

Cell-Cell Communication Analysis

Cell-cell communication (CCC) was calculated using the *CellChat* method[72]. Briefly, *CellChat* utilizes ligand, receptor, and cofactor expression from transcriptomic data to calculate a CCC probability. First, based on a *CellChat*-curated database of ligand-receptor interactions, differentially expressed signaling genes are used to calculate the ensemble average expression of signaling genes. Communication probability is modeled using the law of mass action, and statistically significant communications are identified using a permutation test. We then evaluated the interaction across cell types (excitatory neurons, inhibitory neurons, oligodendrocytes, oligodendrocyte precursor cells, and astrocytes). For the mouse vehicle-treatment comparison, and the human control-AD comparison, the *netVisual_diffInteraction* was used to calculate the net differential interaction strength between each of the two groups.

Cell-Type-Specific Differential Expression Analysis

For each cell type, we performed differential gene expression analysis comparing AD samples to controls using the *FindMarkers* function in the Seurat package. We set the test.use a parameter to MAST, which uses a two-part generalized linear model that models gene expression rate using linear regression and expression level using Gaussian distribution, as recommended by Mou et al[73]. in a study comparing 9 DE methods for single-cell RNA sequencing analysis. We determined differentially expressed genes(DEGs) as those with adjusted p-value < 0.05 based on Bonferroni correction using all genes in the dataset and a log2 fold change (LFC) greater than 0.1.

Pathway Analysis

Pathway enrichment analysis was conducted using g:Profiler, a web tool that performs functional over-representation analysis by mapping a given gene list to known functional information sources and detects statistically significant enriched terms. As input gene lists, we split each cell-type-specific DEG list into upregulated and downregulated subsets and queried them independently with an adjusted p-value cutoff of 0.05 to obtain significant pathway enrichment in both directions. Multiple testing correction was performed using the g:SCS algorithm, which is specifically correct for p-values obtained from GO and pathway enrichment analysis. This method is designed to correct for multiple tests that may be potentially dependent on each other due to term associations. In addition to considering Gene Ontology cellular components, biological processes, and molecular functions as supplementary data, we focused our analysis on the Kyoto Encyclopedia of Genes and Genomes (KEGG) functional pathways.

Computational Drug Repurposing Analysis

The computational drug repurposing algorithm, which was developed by Sirota et al. and Dudley et al., and taken from Chen et al., was applied to each disease gene signature profile using the publicly available Connectivity Map (CMap) database, consisting of treatment profiles of more than 1300 FDA-approved drugs or previously investigated compounds. The drug repurposing algorithm we employed takes two inputs: 1) an ordered list of up and down-regulated genes from individuals with disease as compared to controls and 2) the data from CMap, consisting of rank FC of each gene after drug treatment relative to vehicle controls on the same plate. The pipeline assesses the disease-drug relationship using CMap scores derived from a Kolmogorov-Smirnov test (K-S test), comparing gene expression ranks in disease and by a drug. The pipeline was adapted for single-cell analysis, where each cell-type-specific DEG signature set was overlaid with 6100 drug profiles on the 1300 drugs provided by the CMap. A drug with a strong negative CMap score indicates an opposing mechanistic relationship with the disease, suggesting therapeutic potential by reversing the regulation direction of disease signatures. The absolute values of

CMap scores reflect the degrees to which the drug “flips” the signature of the disease. To address variations in input counts across cell types, significantly reversed drug profiles were identified for each cell type separately using a permutation-based approach. The false discovery rate (FDR; Benjamini-Hochberg) was calculated to adjust P-values. P-values for individual drug hits were determined by comparing reversal scores to a distribution of random scores that were generated by the permutation strategy. Negative reversal scores were deemed significant if they met the criterion of FDR < 0.05. For drugs tested multiple times in different cell lines, the profile with the lowest negative reversal score was used.

Validation in Real-World Human Electronic Medical Records (EMR)

While our drug screening approach generated a total of 86 drug candidates targeting six major brain cell types, we prioritized 25 drugs that significantly reverse more than one brain cell type. This allows maximizing coverage of multiple disease-relevant cell types by combining just two drugs. The beneficial effects of the top drug hits were examined by analyzing AD prevalence (% AD diagnosis) in drug-exposed individuals and compared to propensity-matched controls, via UC-wide EMR. AD diagnosis was defined with ICD10 codes G30.0, G30.1, G30.8, and G30.9. At the time of surveying, the UC-wide EMR aggregated clinical data covering 1,441,778 individuals across six University of California (UC) campuses, encompassing more than 10 million patient records including diagnosis and medication prescriptions. For each screened drug candidate, patients prescribed or taken the drug will be identified using the medication order table and string-matching for drug names. Only individuals above the age of 65 were considered in the study.

The matched controls were identified using a propensity score matching approach based on age, age at death (if applicable), race, ethnicity, sex, original indications (for example, breast cancer was used for Letrozole and colorectal cancer was used for Irinotecan), AD comorbidity (such as hypertension and edema), and UC center locations (including UC San Francisco, UC Davis, UC Los Angeles, UC Irvine, UC San Diego, and UC Riverside). We measured the relative risk score for AD diagnosis (not including patients with only a diagnosis of Mild Cognitive Impairment) in the drug-exposed group and compared it to the matched control groups by calculating the ratios of AD to total patients in each group. Bootstrapped χ-squared tests across 10 permutations of the iterative matching of the control group were applied to establish significance. Drugs with a significantly reduced risk of AD were included in Fig 2.2e. Other drugs found in the UC-wide EMR but did not yield significance or reduced risk were reported in **Table 2.3**. Different original indications were used for matching each drug, and the ICD10 codes were listed in **Table 2.3**.

Drug Selection Rationale for Validation in AD Mouse Model

Our integrated human single-cell transcriptomic analysis revealed two distinct pathological clusters among the six cell types analyzed. Additionally, we observed that AD inhibitory neurons exhibited the most pronounced disparity between AD cases and controls, underscoring its significance in AD pathology. These discoveries informed our prioritization of drug candidates targeting these specific clusters and accommodating inhibitory neurons for treating AD. Among a selection of five promising drug candidates that demonstrated a significant reduction in AD risk in humans, we prioritized letrozole and irinotecan, since letrozole targets the neuronal pathological cluster while irinotecan targets the glial pathological cluster. We hypothesize that the combination of letrozole and irinotecan holds synergistic therapeutic effects, with each pathological cell type cluster addressed by one EMR-validated drug candidate. Consequently, we examined AD outcomes in patients receiving both drugs under their original disease indications in the UC-wide EMR. Unfortunately, fewer than 100 patients were ever prescribed both medications, rendering statistical analysis inconclusive. To explore the synergistic effects of this combination in the context of AD, we turned to direct experimentations in animal models.

Mouse Cohort Generation

The 5xFAD (6SJL) Tg mice (The Jackson Laboratory, ME, USA), overexpress both mutant human amyloid beta (A4) precursor protein 695 (APP) with the Swedish (KM670/671NL), Florida (I716V), and London (V717I) Familial Alzheimer’s Disease (FAD) mutation and human PS1 harboring two FAD mutations, M146L and L286V. These mice were crossed with the widely used tauopathy model, PS19 transgenic mice expressing P301S mutation and including four microtubule-binding domains and one N-terminal insert. A large cohort of the cross, including both sexes, was genotyped, and the litter mates carrying transgenes from both lines were subsequently used for the dosing experiment. Wild-type littermates were kept in the same cages until the behavior test to avoid single-housing of mice.

Drug Treatments

A solution of either letrozole (1 mg/kg) or irinotecan (10 mg/kg) alone, or of the two combined was prepared weekly in 0.9% sterile saline containing 10% Tween-20 to obtain the final concentrations. Treatment dosages were determined with reduced concentrations of tolerable dosage used by previous cancer studies in mice. All treatments were sonicated for 30 minutes before injections to ensure proper mixing of the drugs into the solution. The treatments were given by i.p. injection every other day to the 5xFAD x PS19 double transgenic cohort, evenly split into four groups and sex-balanced, starting 12 weeks before and continuing throughout the behavioral assessment. Mice were aged 4-5 months old before the start of the treatment and the treatments lasted about 4 months until sacrificed. Body weight was measured weekly during drug treatment; injection volume was calculated based on body weight. Control mice were injected with a matched volume of vehicle, 10% Tween in 0.9% sterile saline at pH 8.5. Injections were well tolerated and had no adverse effects on health.

Behavioral Tests

Mice were housed with littermate controls. Each mouse was assigned a random number, so researchers were blinded to genotype and treatment information. Due to the capacity of the test, male and female mice were tested in separate rooms with similar settings and test duration.

The MWM pool (diameter, 122 cm) contained opaque water (22–23°C) with a platform 10 cm in diameter. Mice were trained to locate the platform during three days of “cued” trials, in which the location of the submerged platform is indicated by a black-and-white-striped mast (15 cm high). During these three days, mice received two daily sessions of training, each consisting of two 60s trials with a 15-minute inter-trial interval. Following cued learning, mice were trained for 5 days to locate the hidden platform without the mast. These 5 days also consisted of two training sessions, each consisting of two 60s trials with a 15-minute inter-trial interval. Entry points were changed semi-randomly between trials, but distal visual cues on the walls of the behavioral testing room remained constant throughout the test. To test for memory retention, at 24 and 120 hours after the last hidden platform training, a 60-s probe trial (platform removed) was done. The entry point for the probe trial was in the southwest quadrant, and the target quadrant was the northeast quadrant. Finally, as a control for visual acuity and motor ability, mice were tested with 60s "visible" trials with a visible platform. The platform location remained constant in the hidden platform sessions but was changed for each visible platform session. Performance was monitored with an EthoVision video-tracking system (Noldus Information Technology). For the probe trials, we analyzed (1) target quadrant preference- the percent time spent in the target quadrant versus average time spent in the three other quadrants, (2) platform crossing- the number of crossings over the position of the target platform versus the average number of crossings over the equivalent positions in the three other quadrants as well as (3) escape latency- a memory probe measuring how fast mice arrive at the platform location the first time after placement in the maze. An ANOVA was used to analyze the effect of treatment, genotype, and probe timing on percent on-target crosses.

Histopathological Analyses

One hemibrain per mouse was drop-fixed in 4% paraformaldehyde prepared in 1xPBS, washed for 24 hours in 1xPBS, then cryoprotected in 30% sucrose for 48 hours at 4°C. The fixed hemispheres were cut into 30 μm thick coronal slices using a freeze-sliding microtome (Leica). These sections were then stored in a cryoprotectant solution of 30% ethylene glycol, 30% glycerol, and 40% 1xPBS at −20 °C.

Hippocampal brain sections (ten sections per mouse spaced approximately 300 µm apart, 30-µm thick) were mounted onto microscope slides from Fisher Scientific. A 0.1% Sudan Black solution was prepared by dissolving Sudan Black powder (Sigma) in 70% ethanol (KOPTEC) and mixing it with a magnetic stirrer. After centrifuging the solution at 1,100g for 10 minutes, the supernatant was filtered using a 0.2-µm filter syringe (Thermo Scientific). The brain sections were stained with the 0.1% Sudan Black solution for 10 minutes, then washed in 70% ethanol followed by Milli-Q water. The sections were then coverslipped with ProLong Gold mounting medium (Invitrogen). All sections were imaged; eight consecutive brain sections with consistent anatomic locations across mice were quantified.

For Thio-S staining, several brain sections spaced 300 µm apart were mounted onto slides, following a protocol adapted from a previous study. The sections were washed with 1× PBS-T and then incubated in a solution of 0.06% Thio-S in PBS for 8 minutes. After this incubation, the sections were washed for 1 minute in 80% ethanol and then for 5 minutes in PBS-T. The sections were counterstained with DAPI for 8 minutes, washed again with PBS-T, and coverslipped. Three sections per mouse were quantified and averaged for percent Thio-S positive area and number of plaque counts (normalized by hippocampal area size per section).

For immunofluorescent staining, several brain sections spaced approximately 300 µm apart, were washed three times with 1× PBS-T (PBS + 0.1% Tween-20) (Millipore Sigma) and then incubated for 5 minutes in boiling antigen retrieval buffer (Tris buffer, pH 7.6; TEKNOVA). After this, the sections were rinsed in PBS-T before being placed in a blocking solution composed of 5% normal donkey serum (Jackson Labs) and 0.2% Triton-X (Millipore Sigma) in 1× PBS for 1 hour at room temperature. Next, the sections were washed again in PBS-T and incubated in Mouse-on-Mouse (MOM) Blocking Buffer (one drop MOM IgG in 4 ml PBS-T) (Vector Labs) for another hour at room temperature. Following the MOM block, the sections were incubated overnight at 4°C in primary antibodies diluted to their optimal concentrations. The antibodies and their dilutions included: anti-AT8 (ms, 1:300, Invitrogen), anti-GFAP (ms, 1:500, Millipore

Sigma), anti-Iba1 (rbt, 1:300, Wako), anti-NeuN (GP, 1:500, Millipore Sigma). Three sections per mouse were quantified and averaged for a positive percent area.

After the primary antibody incubation, the sections were washed in PBS-T and then incubated for 1 hour at room temperature in secondary antibodies. This included donkey anti-mouse 488 (1:1,000, Abcam), donkey anti-rabbit 594 (1:1,000, Abcam), donkey anti-guinea pig 594 (1:1,000, Jackson Immuno), and donkey anti-guinea pig 647 (1:1,000, Jackson Immuno). Subsequently, the sections were washed in PBS-T and incubated in DAPI (1:30,000 dilution in PBS-T) (Thermo Fisher) for 8 minutes at room temperature. After a final wash with PBS-T, the sections were mounted onto microscope slides (Fisher Scientific), coverslipped with ProLong Gold mounting medium (Vector Laboratories), and sealed with clear nail polish.

Images for quantifications were captured using a scanning microscope (Keyence) at magnifications of ×10 or ×20, depending on the stain. To minimize batch-to-batch variation, all samples for each stain were processed simultaneously and imaged at the same fluorescent intensity. For quantifying the percent coverage area, an optimal threshold was established for each stain in ImageJ, and all samples were quantified using this threshold. To prevent bias, researchers were blinded to the sample identities. Representative images were captured using an Aperio VERSA slide scanning microscope (Leica) at ×20 magnification.

snRNA-Seq Library Preparation and Sequencing

The other hemibrains were dissected by brain subregions, rapidly frozen on dry ice, and kept at −80 °C. Hippocampal samples were used for single-nuclei preparation. One frozen mouse hippocampus was placed into a pre-chilled 2 mL Dounce with 1 mL of cold 1X Homogenization Buffer (1X HB) (250 mM Sucrose, 25 mM KCL, 5 mM $MgCl_2$, 20 mM Tricine-KOH pH7.8, 1 mM DTT, 0.5 mM Sermidine, 0.15 mM Sermine, 0.3% NP40, 0.2 units/µL RNase inhibitor, ~0.07 tabs/ml cOmplete Protease inhibitor). Dounce with “A” loose pestle (~10 strokes) and then with “B” tight pestle (~10 strokes). The homogenate was filtered using a 70 µM Flowmi strainer (Bel-Art) and transferred to a pre-chilled 2 mL LoBind tube (Fischer Scientific). Nuclei were pelleted by spinning for 5 min at 4°C at 350 RCF. The supernatant was removed and the nuclei were resuspended in 400 µL 1X HB. Next, 400 µL of 50% Iodixanol solution was added to the nuclei and then slowly layered with 600 µL of 30% Iodixanol solution under the 25% mixture, then layered with 600 µL of 40% Iodixanol solution under the 30% mixture. The nuclei were then spun for 20 min at 4°C at 3,000RCF in a pre-chilled swinging bucket centrifuge. 200 µL of the nuclei band at the 30%-40% interface was collected and transferred to a fresh tube. Then, 800 µL of 2.5% BSA in PBS plus 0.2 units/µL of RNase inhibitor was added to the nuclei and then were spun for 10 min at 500 RCF at 4C. The nuclei were resuspended with 2% BSA in PBS plus 0.2 units/µL RNase inhibitor to reach at least 500 nuclei/µL. The nuclei were then filtered with a 40 µM Flowmi cell strainer. The nuclei were counted and then ~13,000 nuclei per sample were loaded onto 10x Genomics Next GEM chip M. The snRNA-seq libraries were prepared using the Chromium Next GEM Single Cell 3′ HT kit v3.1 (10x Genomics) according to the manufacturer’s instructions. Libraries were sequenced on an Illumina NovaSeq 6000 sequencer at the UCSF CAT sequencing core.

Sequence Alignment, Filtering, and Counting

The demultiplexed fastq files were processed following the procedure previously described by Zalocusky et al. In summary, the fastq files were aligned to the mouse reference genome, mm10-1.2.0, which includes introns, using the *cellranger count* function (version 4.0.0) with default parameters, as detailed in the Cell Ranger documentation. Subsequently, a single UMI count file per animal/sample was generated by the *cellranger count* function. Individual UMI count files were then combined into a single count matrix using the *merge* function in the Seurat package v4.0.4. Metadata, including age, sex, and treatment information, were added to each cell.

Pre-Processing and Quality Control

The count matrix was further processed with Seurat by first calculating the percentage of mitochondria genes mapped per cell. The distribution of feature count, total mapped gene count, and percentage of mitochondria genes were visualized across biological samples as violin plots, and no obvious outlier was identified. We filtered the count matrix to only include cells with higher than 250 gene features, at least 500 gene counts, and mitochondria gene percentages lower than 10%. Potential misaligned or ambiguous gene features expressed in fewer than 10 cells were also removed. These quality assurance steps resulted in a final Seurat object containing 25,642 gene features expressed by 223,546 nuclei.

Normalization, Dimensional Reduction, and Clustering

Count normalization and dimensionality reduction were conducted following standard procedure in the Seurat package. In brief, we performed normalization and variance stabilization with an updated version of *sctransform*, v2, and principal component analysis (PCA) with *RunPCA* (npcs = 30). Dimensional reduction through Uniform Manifold Approximation and Projection (UMAP) was performed with the *RunUMAP* function and considering the top 15 dimensions selected from the corresponding PCA.

The clustering was based on the first 15 principal components (PCs) using the *FindNeighbors* function in Seurat. This function embeds cells in a K-nearest neighbor graph, considering the Euclidean distance in PCA space and refining the edge weights between any two cells based on the shared overlap in their local neighborhoods. The clustering was obtained using the *FindClusters* function, which employs modularity optimization techniques such as the Louvain algorithm, with a resolution parameter of, resulting in a set of 35 distinct clusters.

Mouse Cell Type Annotation, Differential Gene Expression, and Pathway Enrichment Analysis

Major cell types, including astrocytes, microglia, oligodendrocytes, oligodendrocyte precursor cells, and excitatory and inhibitory neurons, were classified using mouse brain cell markers in PanglaoDB, a publicly available marker gene database. Further subdivisions of hippocampus cell types, such as CA1 and CA3 pyramidal cells, were queried against hippocampal cell-type-specific marker genes as published in hipposeq (https://hipposeq.janelia.org). Cell type identities per cluster were determined by applying Seurat's *AddModuleScore* function to sets of mouse brain marker genes. A module score for each cell type considered was calculated per cell. Each cell was assigned the corresponding cell type identity that generated the highest scores among scores for all cell types. If the highest and second highest scores of a cell were within 20% of the highest score, then the cells were deemed hybrids and excluded from further analysis. We assessed the validity of the assigned cell type identities by examining the homogeneity, distribution, and separation of cell types by clustering in UMAP plots. Minority cell types in a cluster, defined as a cell type that accounts for less than 5% of the total counts for that cluster, were considered potential hybrid cells and excluded from further analysis.

For the cell-type-specific differential gene expression analysis, we used the same procedure and tools in the human-integrated analysis. For pathway enrichment analysis, in addition to using g:Profiler, we curated the EnrichR-KG web tool, which facilitates analysis across multiple databases and provides visual representations linking significantly enriched genes with the associated pathways and GO terms. The reversal gene-pathway network was generated and downloaded using EnrichR-KG with significantly differentially reversed genes as inputs. For each cell-type-stratified analysis, the top five KEGG pathways and GO terms were displayed in a network format

2.4 Figures

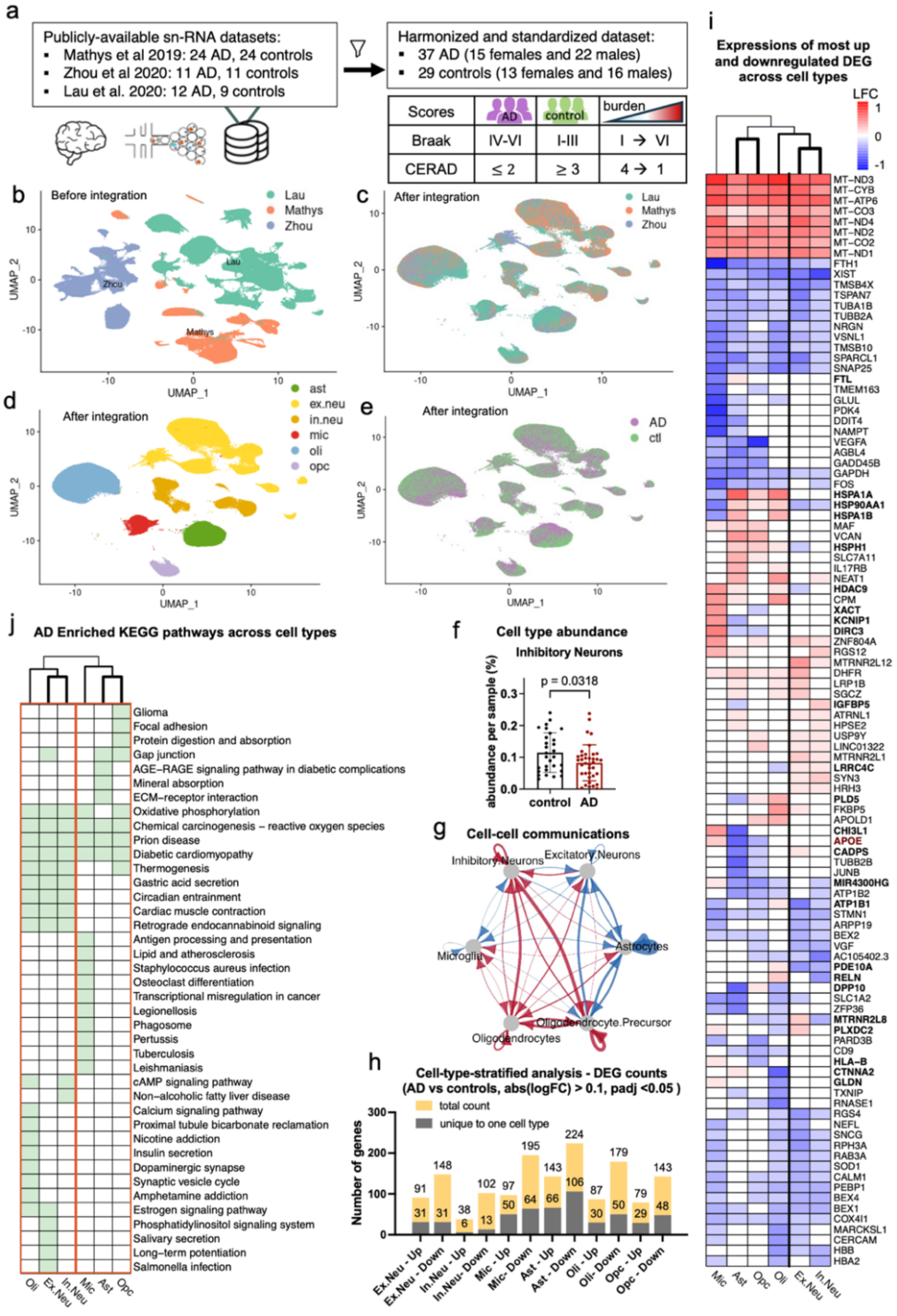
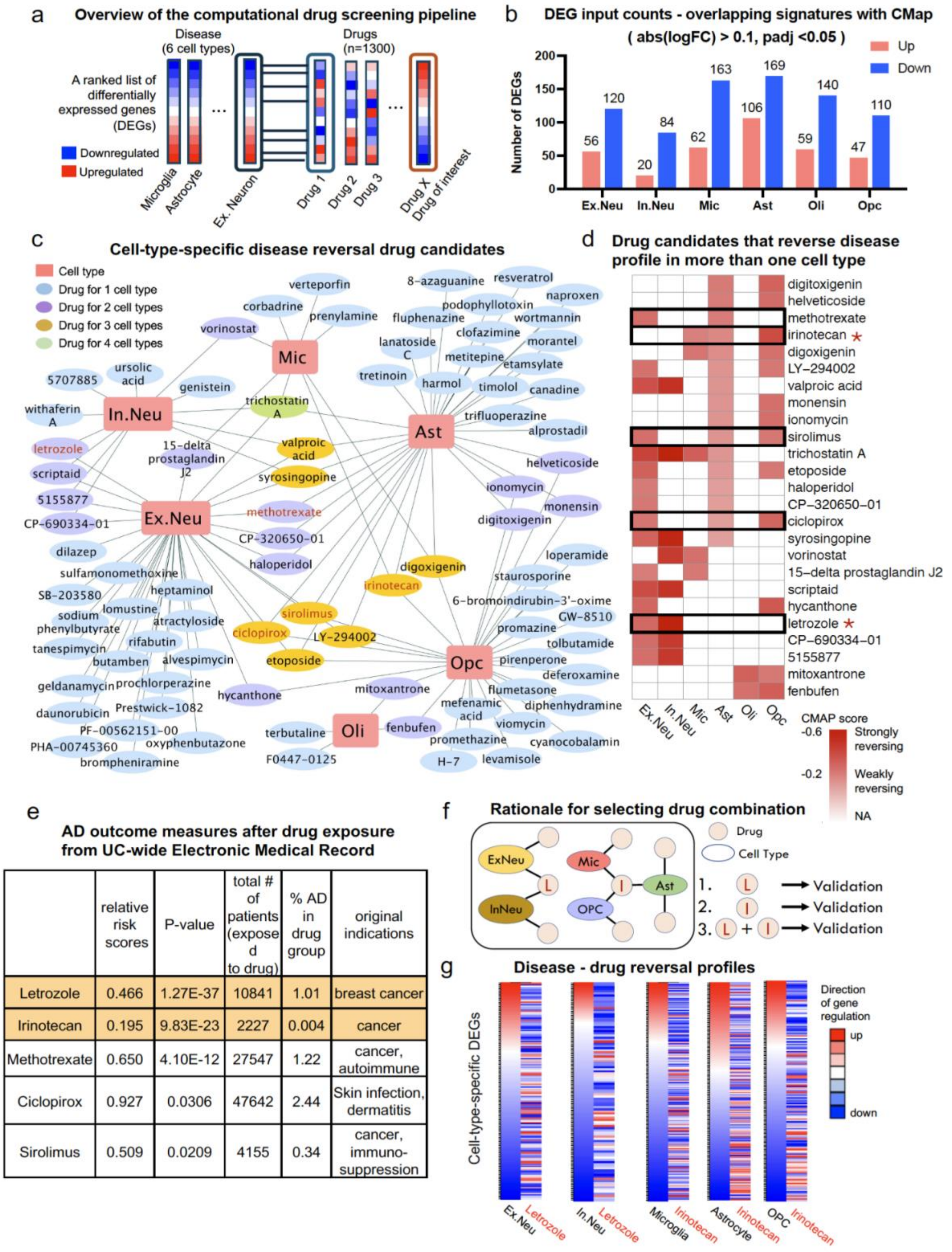


Figure 2.4.1 Single nucleus transcriptomic profiling reveals both shared and cell-type-specific gene expression signatures in human AD samples.

Figure 2.4.1 Single nucleus transcriptomic profiling reveals both shared and cell-type-specific gene expression signatures in human AD samples. a. Data sources (public-available) and cohort summary, including sample filtering criteria and case-control standardization strategies. b. UMAP plot of merged samples from three independent studies. Study-associated variations are apparent without integration. c-e. UMAP plots of harmonized dataset using Seurat CCA integration algorithm, with study variation eliminated, and labeled by study (c), cell type identity (d), and case-control identification (e). f. Scattered bar plot of cell type abundance percentage for inhibitory neurons, comparing AD and controls (unpaired two-sided t-test). g. Projection plots for cell type to cell type communication measured by expression of receptor and ligand pairs and represented as arrows connecting two cell types. Red connections indicate increased communication and blue connections indicate decreased communication comparing AD vs controls. Arrows point to the directions from the senders to the receivers of the communication. h. Bar plot of AD versus controls differentially expressed gene (DEG) counts in the six major brain cell types and separated into up and down-regulated subgroups. Counts of DEGs unique to one cell type were superimposed on the total counts. i. DEG heatmap showing overlaps of top DEGs (highest absolute log fold changes) in each cell type, DEGs with opposite regulatory patterns in AD across cell types are bolded. j. Heatmap showing AD-enriched KEGG pathways across cell types. Few functional pathways are shared by all six cell types. Oligodendrocytes cluster with excitatory and inhibitory neurons, suggesting more similarity among these cell types, while microglia cluster with astrocytes and OPCs.



	relative risk scores	P-value	total # of patients (exposed to drug)	% AD in drug group	original indications
Letrozole	0.466	1.27E-37	10841	1.01	breast cancer
Irinotecan	0.195	9.83E-23	2227	0.004	cancer
Methotrexate	0.650	4.10E-12	27547	1.22	cancer, autoimmune
Ciclopirox	0.927	0.0306	47642	2.44	Skin infection, dermatitis
Sirolimus	0.509	0.0209	4155	0.34	cancer, immuno-suppression

Figure 2.4.2 Computational drug repurposing pipeline predicts drug candidates reversing cell-type-specific transcriptomic signature profiles of AD.

Figure 2.4.2 Computational drug repurposing pipeline predicts drug candidates reversing cell-type-specific transcriptomic signature profiles of AD. a, Schematic showing computational drug repurposing workflow. b, Input DEG counts (filtered list by mapping AD vs control DEGs to existing gene probes in the CMap database) per cell type for the computational drug repurposing pipeline. c, Disease-drug network depicting the connections between the six cell types and drug candidates that significantly reverse disease profiles within the respective cell types. Drug names in red denote drug candidates validated in humans using electronic medical records (EMR). d. Heatmap showing drug candidates that significantly reverse AD profiles in more than one cell type. Black frames label drugs validated in humans using the EMR, and red asteroids label drugs selected for validation *in vivo* using a mouse model of AD. e, AD prevalence assessment in drug-exposed individuals, only drugs with significantly reduced AD risks are shown. f, Schematic illustration demonstrating the rationale of prioritizing letrozole and irinotecan as potential combination therapy for further validation in a mouse model of AD. g, Heatmaps of each cell-type-specific AD transcriptomic signature profile, rank ordered genes from the most upregulated to the most downregulated and color-coded by log fold changes, in comparisons with the gene probe ranks by letrozole or irinotecan treatments, colored by corresponding fold change ranks in the CMap.

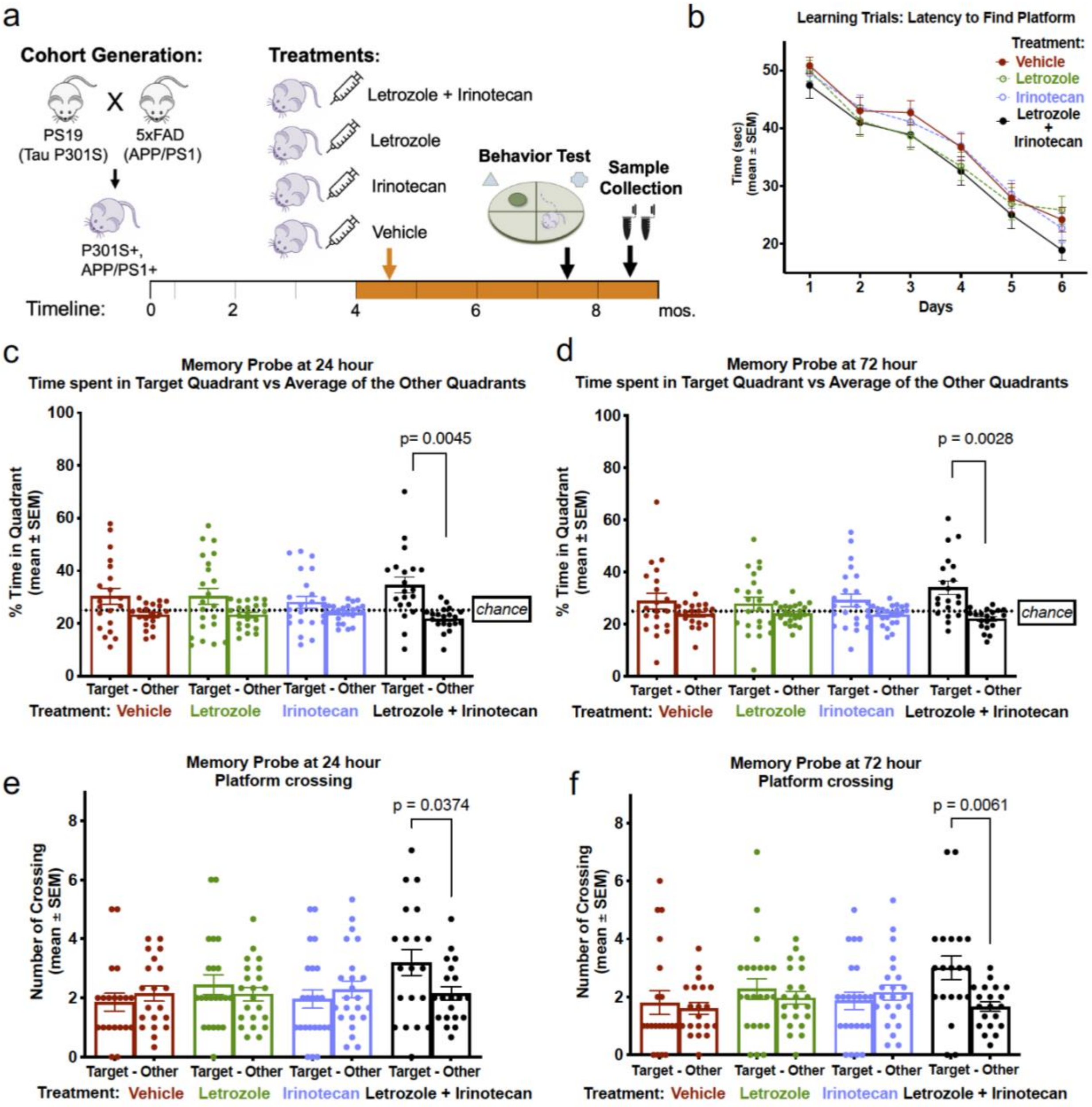


Figure 2.4.3 Combination treatment with letrozole and irinotecan rescues AD-like memory impairments in aged 5xFAD/PS19 mice with both Ab and tau pathologies.

a. The schematic of treatment cohort design and experimental timeline. Four treatment groups (n=20 per group, both sexes), including treatment with vehicle, letrozole (1mg/kg), irinotecan (10mg/kg) or combination of both drugs every other day via i.p. injection. b. Escape latency over learning days 1–6 did not differ statistically between groups. One-way repeated-measures ANOVA test was applied comparing vehicle group with treatment groups. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.) c,d. Memory probes of percent time spent in the target quadrant versus the average of other quadrants demonstrated a significant preference of the target quadrant solely by mice with combination treatment at 24 hour (c) and 72 hour (d) after removing the hidden platform. e,f. Memory probes measuring numbers of platform-location crossing in the target quadrant versus average of the other quadrants. Significantly more crossings in the target quadrant where the platform used to be only observed in the combination treatment group at 24 hour (e) and 72 hour (d) after the platform was removed. c-f. Ordinary one-way ANOVA test was performed between the target and other quadrants for each group, and p-values were adjusted with Bonferroni multiple-comparisons testing.

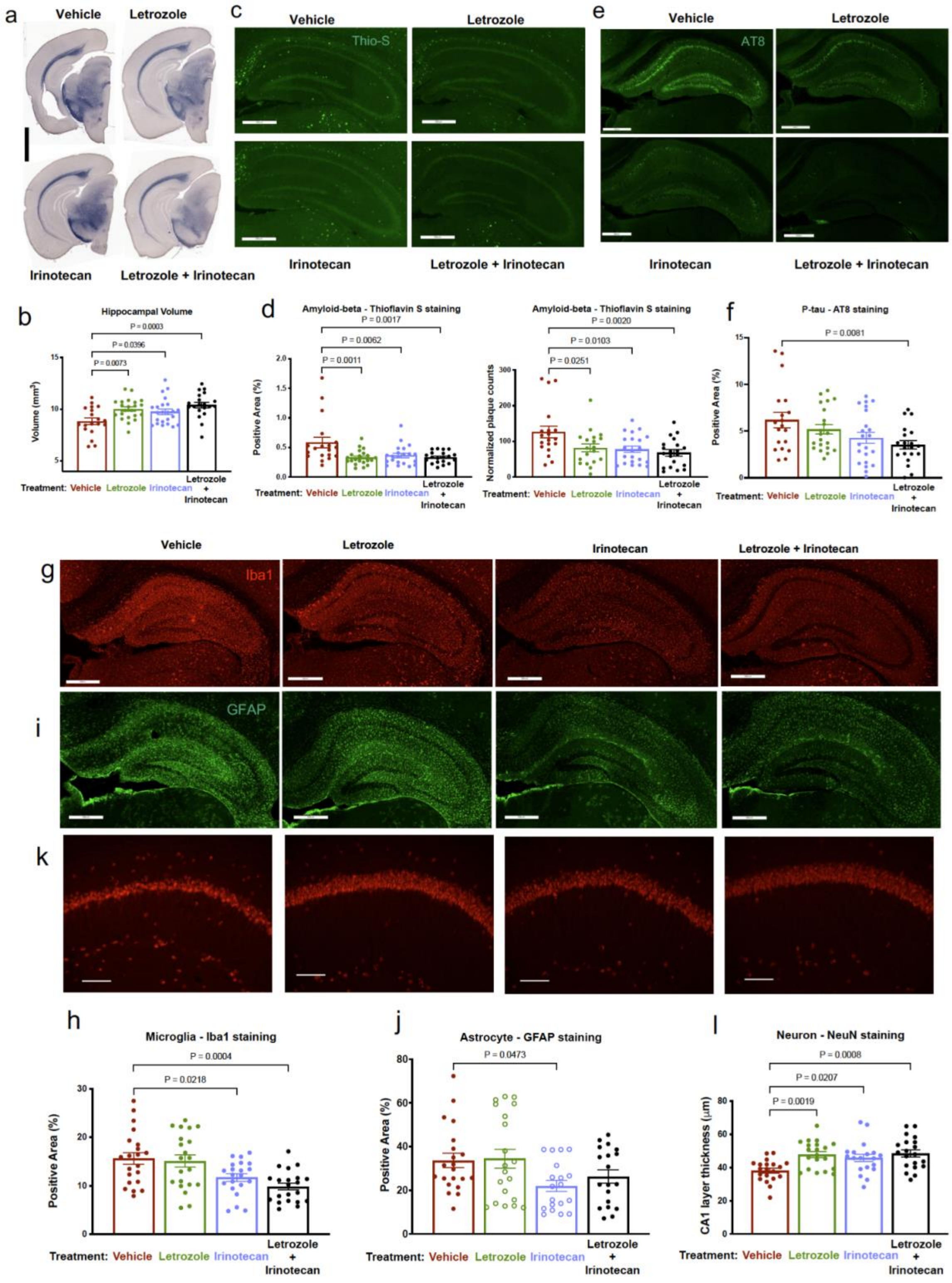


Figure 2.4.4 AD pathologies are significantly reduced in 9-month-old 5xFAD/PS19 mice after drug treatments, with the strongest rescue in the combination treatment group.

Figure 2.4.4 AD pathologies are significantly reduced in 9-month-old 5xFAD/PS19 mice after drug treatments, with the strongest rescue in the combination treatment group. a. Representative images of the ventral hippocampus from 9-month-old 5xFAD/PS19 mice with Sudan Black staining to enhance hippocampal visualization (scale bar, 2mm). b. Quantification of hippocampal volume in 9-month-old 5xFAD/PS19 mice across treatment groups. c. Representative images of hippocampus from 9-month-old 5xFAD/PS19 mice with immunostaining of phosphorylated tau (p-tau) using AT8 monoclonal antibody (scale bar, 500mm). d. Quantification of AT8-positive p-tau percent coverage area in 9-month-old 5xFAD/PS19 mice across treatment groups. e. Representative images of Thio-S staining in the hippocampus from 9-month-old 5xFAD/PS19 mice (scale bar, 500mm). f. Quantification of Thio-S-positive percent coverage area and plaque counts in the hippocampus of 9-month-old 5xFAD/PS19 mice across treatment groups. g. Representative images of microglia immunostaining with anti-Iba1 in the hippocampus of 9-month-old 5xFAD/PS19 mice (scale bar, 500mm). h. Quantification of the percent Iba1 coverage area in the hippocampus of 9-month-old 5xFAD/PS19 mice across treatment groups. i. Representative images of astrocyte immunostaining with anti-GFAP in the hippocampus of 9-month-old 5xFAD/PS19 mice (scale bar, 500mm). j. Quantification of percent GFAP coverage area in hippocampus of 9-month-old 5xFAD/PS19 mice. k. Representative images of CA1 neurons with neuronal marker NeuN immunostaining (scale bar, 100mm). l. Quantification of the thickness of the CA1 neuronal cell layer of 9-month-old 5xFAD/PS19 mice across treatment groups. b, d, f, h, j, l. Ordinary one-way ANOVA were applied between all treatment groups and vehicle control group, p-values were adjusted by Dunnett's multiple comparison test

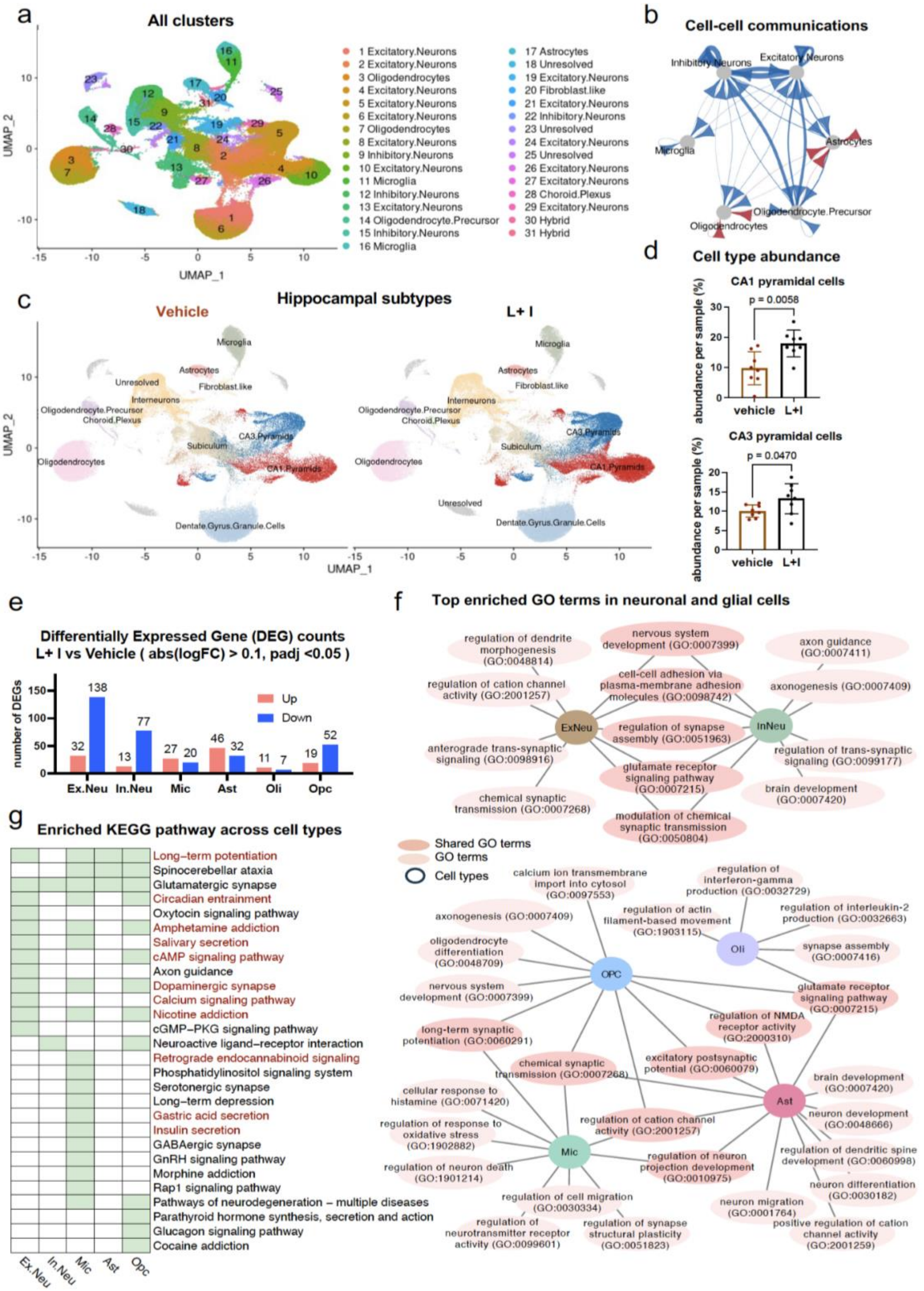


Figure 2.4.5 Single-nucleus RNA-sequencing (snRNA-seq) analysis in 9-month-old 5xFAD/PS19 mice across treatment groups.

Figure 2.4.5 Single-nucleus RNA-sequencing (snRNA-seq) analysis in 9-month-old 5xFAD/PS19 mice across treatment groups. a. UMAP plot of all 31 distinct cell clusters in hippocampus of mice from combination-treatment and vehicle-treatment groups. b. Projection plots for cell type to cell type communication measured by expression of receptor and ligand pairs and represented as arrows connecting two cell types. Red connections indicate increased communication and blue indicate decreased communications comparing L+I-treated versus vehicle-treated groups. Arrows points to the directions from sender to receiver of the communication. c. UMAP plots labeled by hippocampal specific cell types and split by treatments to highlight difference in distribution density for neuronal subtypes between treatments. d. Scattered bar plots of cell type abundance (percentage) for CA1 and CA3 pyramidal neurons (unpaired two-sided t-test), comparing L+I-treated and vehicle-treated mice (n=8 per group, including both sexes). e. Differentially expressed gene (DEG) counts of L+I-treated versus vehicle-treated mice in six major brain cell types. Counts of DEGs per cell type do not correlate with cell counts. f. Top combination treatment-enriched Gene Ontology (GO) terms across six cell types. g. Heatmap illustration of enriched KEGG pathways across six cell types. Some treatment-enriched pathways overlap with AD signature pathways from integrated human snRNA-seq analysis (see Figure 2.1) and were labeled in red.

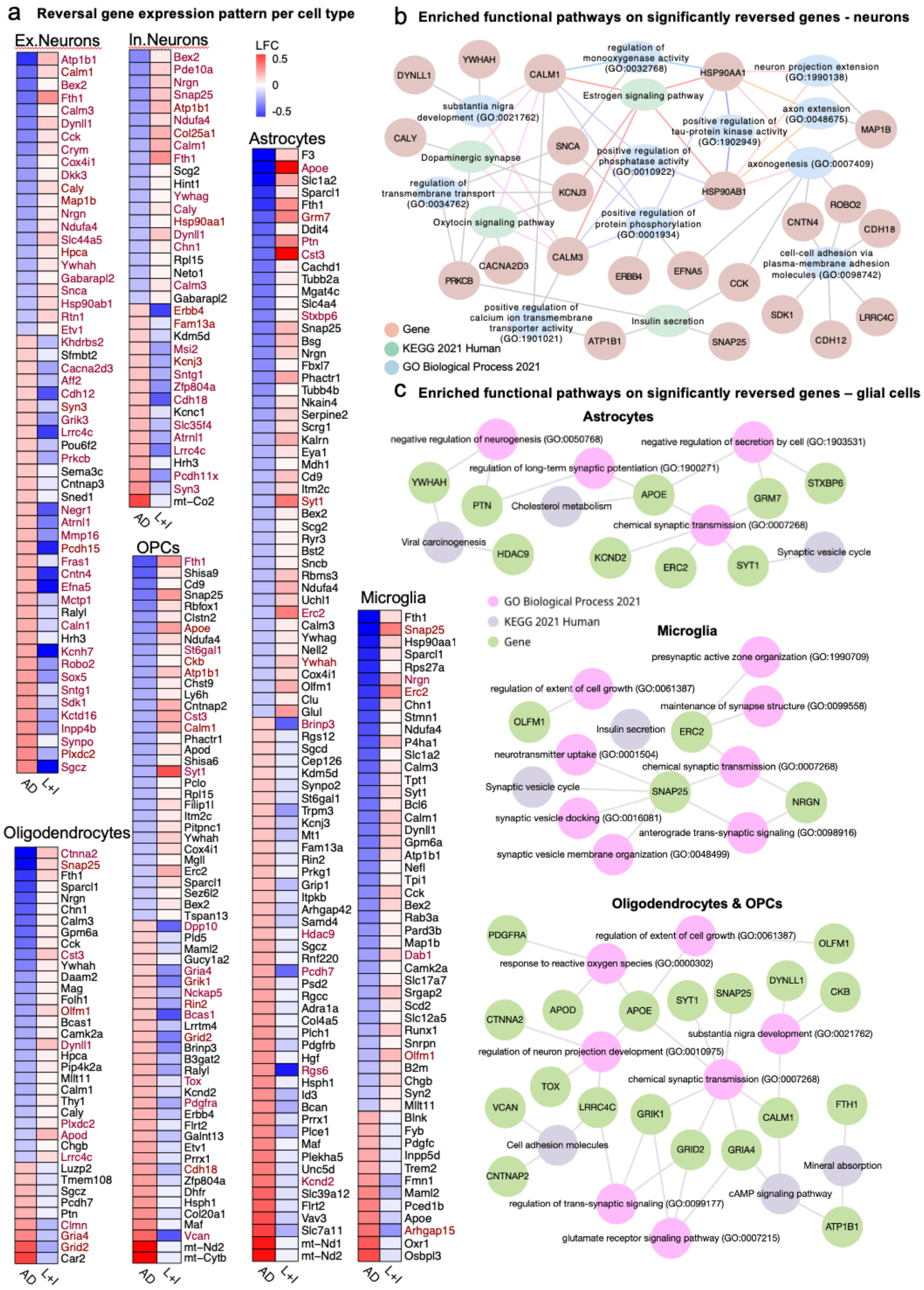


Figure 2.4.6 Combination treatment with letrozole and irinotecan reverses cell-type-specific transcriptomic signatures of AD.

Figure 2.4.6 Combination treatment with letrozole and irinotecan reverses cell-type-specific transcriptomic signatures of AD. a. Comparison of cell-type-specific transcriptomic signature profiles of AD in humans with gene expression changes in combination treatment versus vehicle treatment groups of mice. Only treatment-reversed genes from AD profiles are included (abs (LFC) >0.01 in L+I-treated versus vehicle-treated mice), and those with FDR-adjusted p value < 0.05 are labeled in red. Heatmap colors indicate directions and magnitude of gene expression changes, with downregulations in blue and upregulations in red. b. Gene set enrichment analysis on treatment-reversed genes, which reached statistical significance, in excitatory and inhibitory neurons revealed reversal of multiple AD-relevant functional pathways. c. Gene set enrichment analysis on treatment-reversed genes, which reached statistical significance, in astrocytes, microglia, or oligodendrocyte and OPCs demonstrated reversal of glial cell specific AD gene signatures and functional pathways.

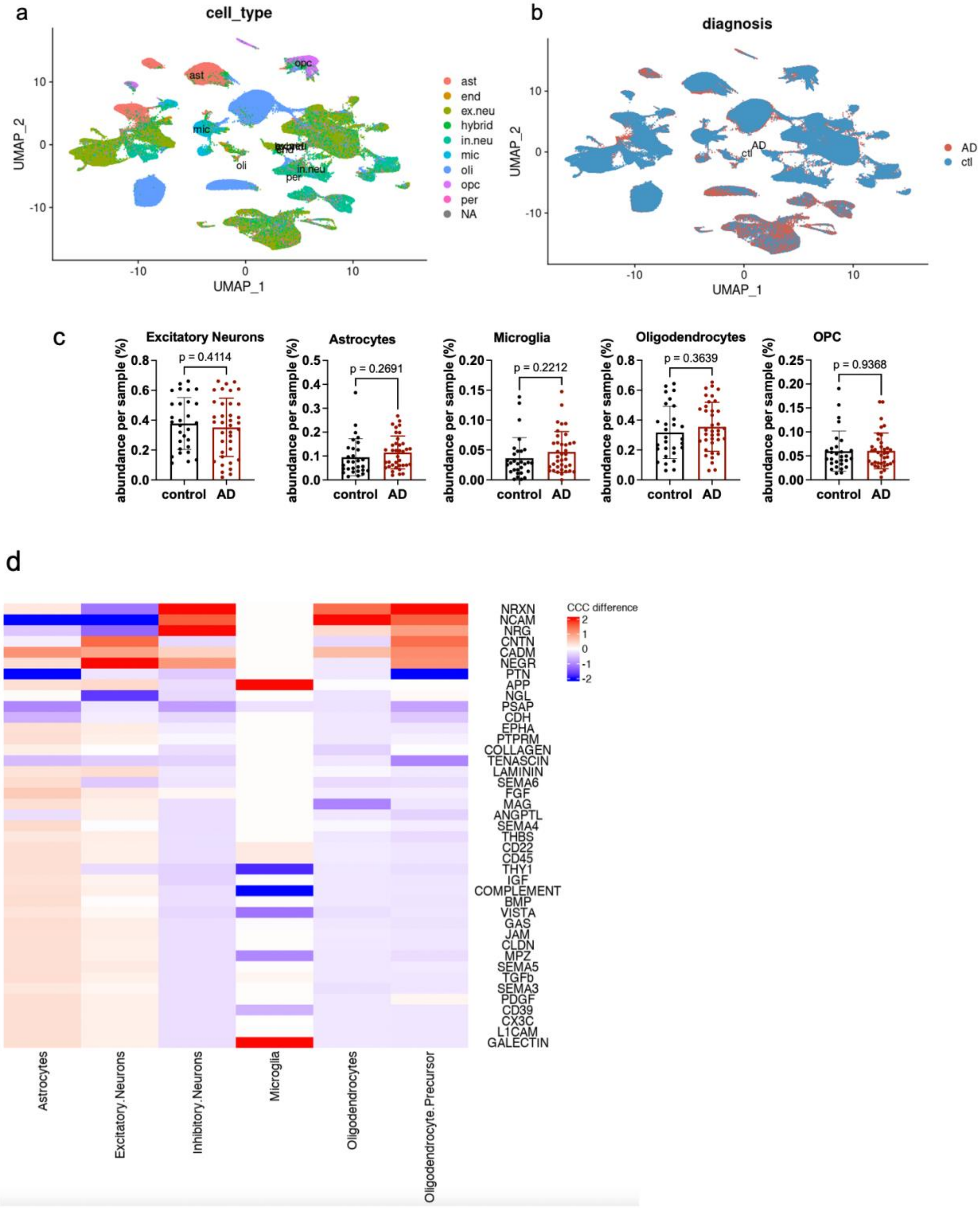


Figure 2.4.7 Extended data on human integrated dataset

Figure 2.4.7 Extended data on human integrated dataset a, b. UMAP plot of merged samples before CCA integration, labeled with cell type identities (a) and disease status (b). c. Scattered plots of cell type abundance (percentage) for the other five cell types, comparing AD and controls (unpaired two-sided t-test). d. Heatmap displaying pathway-level cell-cell-communication (CCC) difference comparing AD and controls and stratified by cell type. Color value represents the scaled CCC probability difference in AD compared to controls. Signaling pathways represent ligand-receptor pairs based on the KEGG Pathway database.

Figure 2.4.8 Extended data on human integrated dataset

Figure 2.4.8 Extended data on human integrated dataset a. UpSet plot showing Gene Ontology (GO) term counts for the six major brain cell types and intersects across cell types. b. Heatmap showing shared AD-enriched GO terms enriched in the six cell types. c. Heatmap showing AD-enriched GO terms unique to each cell type.

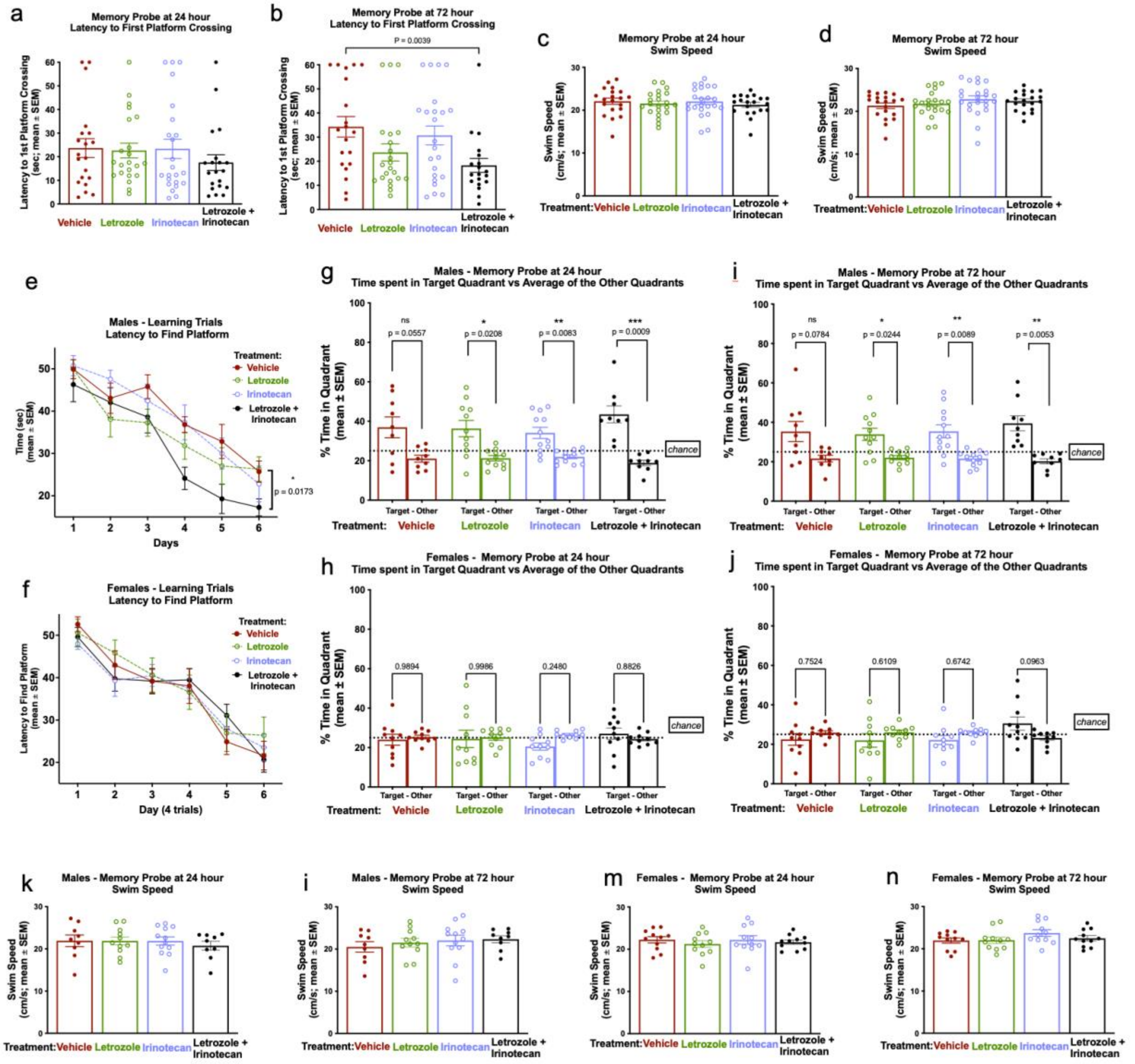


Figure 2.4.9 Extended data on behavior tests

a,b. Memory probes measuring latency to first time of platform crossing at 24 hour(a) and 72 hour (b) demonstrated long-term location memory rescue in combination treated mice only, at 72 hour after platform removed. c,d. Swim speed at 24hour (a) and 72 hour (b), no significant difference in swim speed across groups indicating that the observed behavioral differences were not confounded by physical performance or motor function. e,f, Escape latency over learning days 1-6 in male mice (e) and in female mice (d) demonstrating significant better learning in combination-treated males. One-way repeated-measures ANOVA test was applied comparing vehicle group with treatment groups with Sidak's multiple comparison test. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.) g-j. Memory probes of percent time spent in target quadrant vs average of other quadrant demonstrating a significant preference of the target quadrant by letrozole, irinotecan, and combination-treated males but not the vehicle-treated males at 24 hour (g) and 72 hour (i) after removing the hidden platform. No significant preference was observed in females at 24 hour (h) and 72 hour (j). Ordinary one-way ANOVA test was performed between target and other quadrants for each group, and p-value adjusted with Bonferroni multiple-comparisons testing. k-n. No difference in swim speeds was observed in males at 24 hour (k) and 72 hour (i) or females at 24 hour (m) and 72 hour (n). c,d,k,i,m,n. Ordinary one-way ANOVA were applied between all treatment groups and vehicle control group, p-values were adjusted by Dunnett’s multiple comparison test.

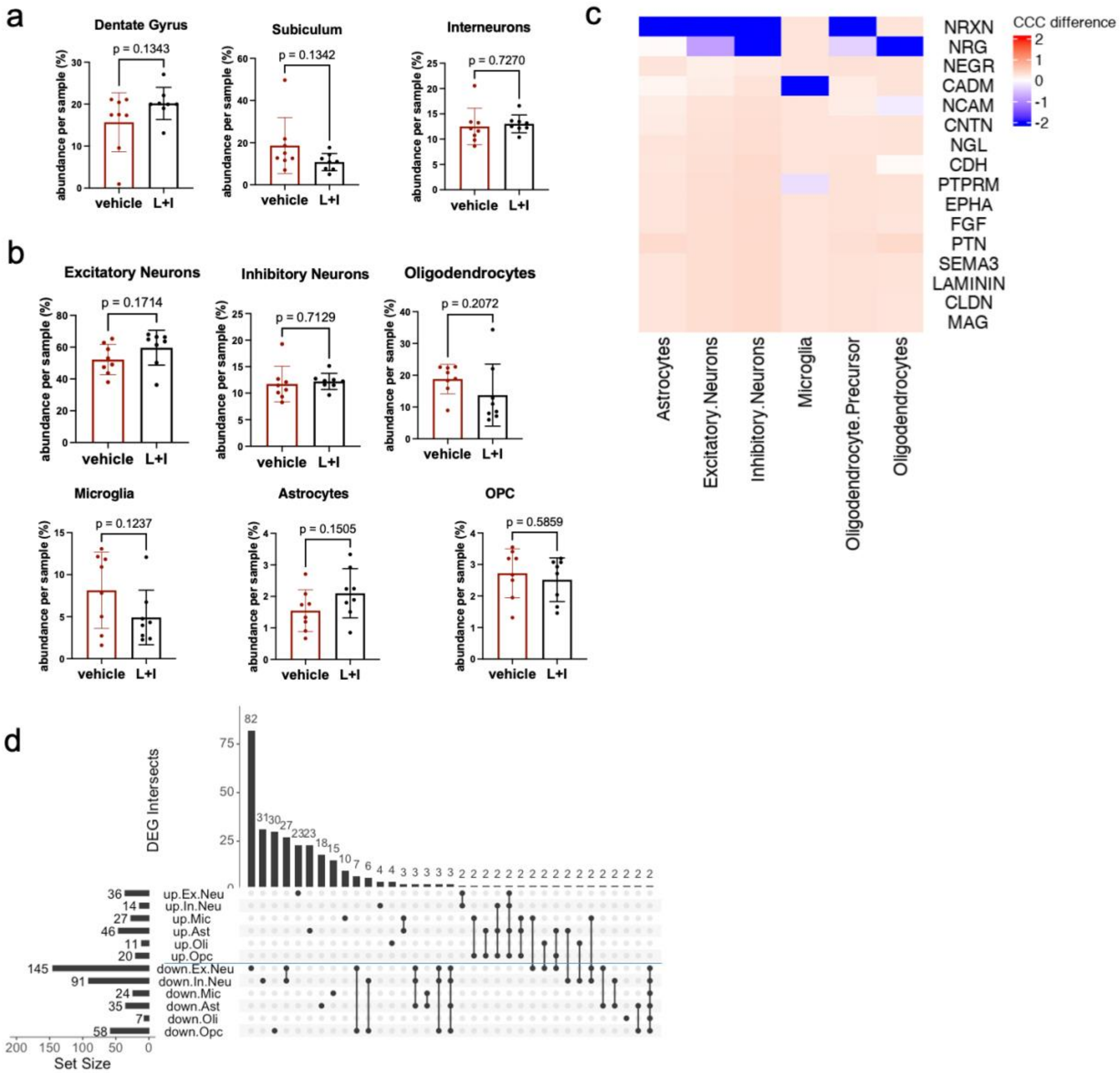


Figure 2.4.10 Extended data on treatment dataset

a. Cell type abundance analysis for neuronal subtypes of hippocampus. b. Cell type abundance analysis for the six major brain cell types. c. Heatmap displaying pathway-level Cell-cell-communication difference between combination-treated and vehicle-treated and stratified by cell type. Value represents the scaled CCC probability difference in combination-treated compared to vehicle-treated. Signaling pathways represent ligand-receptor pairs based on the KEGG Pathway database. d. UpSet plot showing DEG counts (combination-treated compared to vehicle-treated, abs(LFC) > 0.1 & p-adj-value < 0.05) and intersect across groups (separated by cell types and direction of regulation).

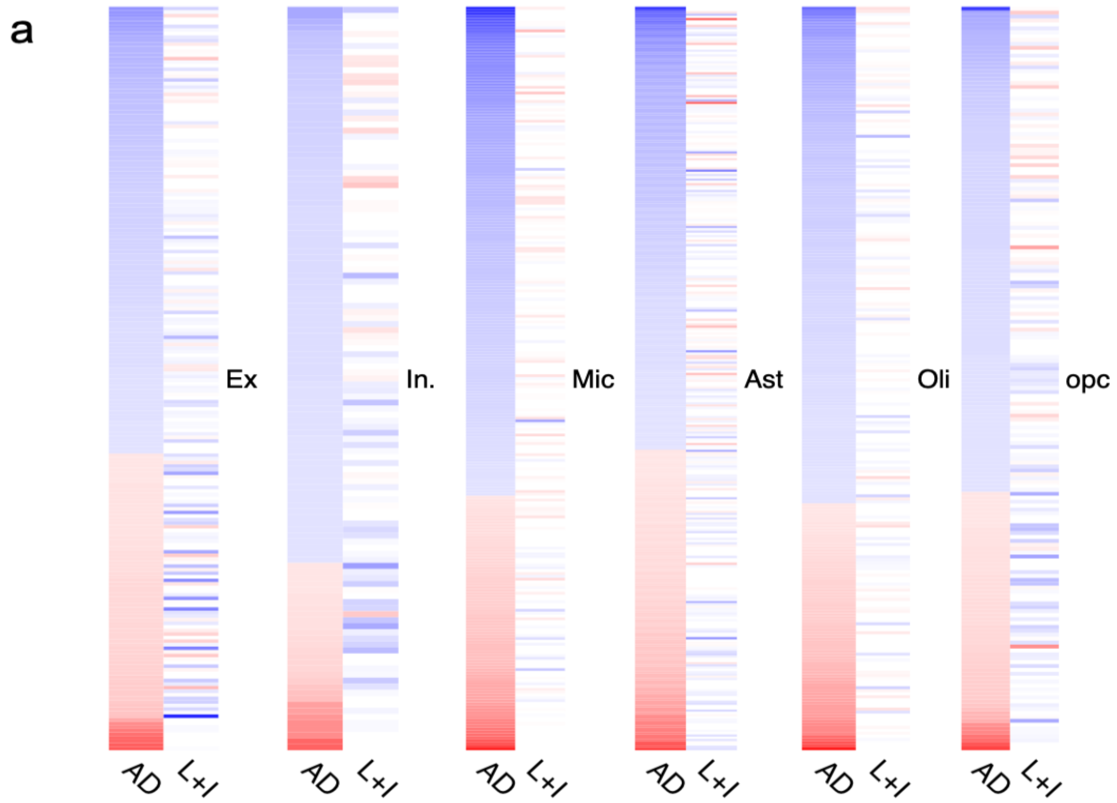


Figure 2.4.11 Extended data on gene reversal across cell types

a. All AD signature gene expression pattern comparison in human and after L+I treatment. Red indicates upregulation while blue indicates downregulation. White labels gene with no expression difference in treatment vs control or not expressed in mice.

2.5 Tables

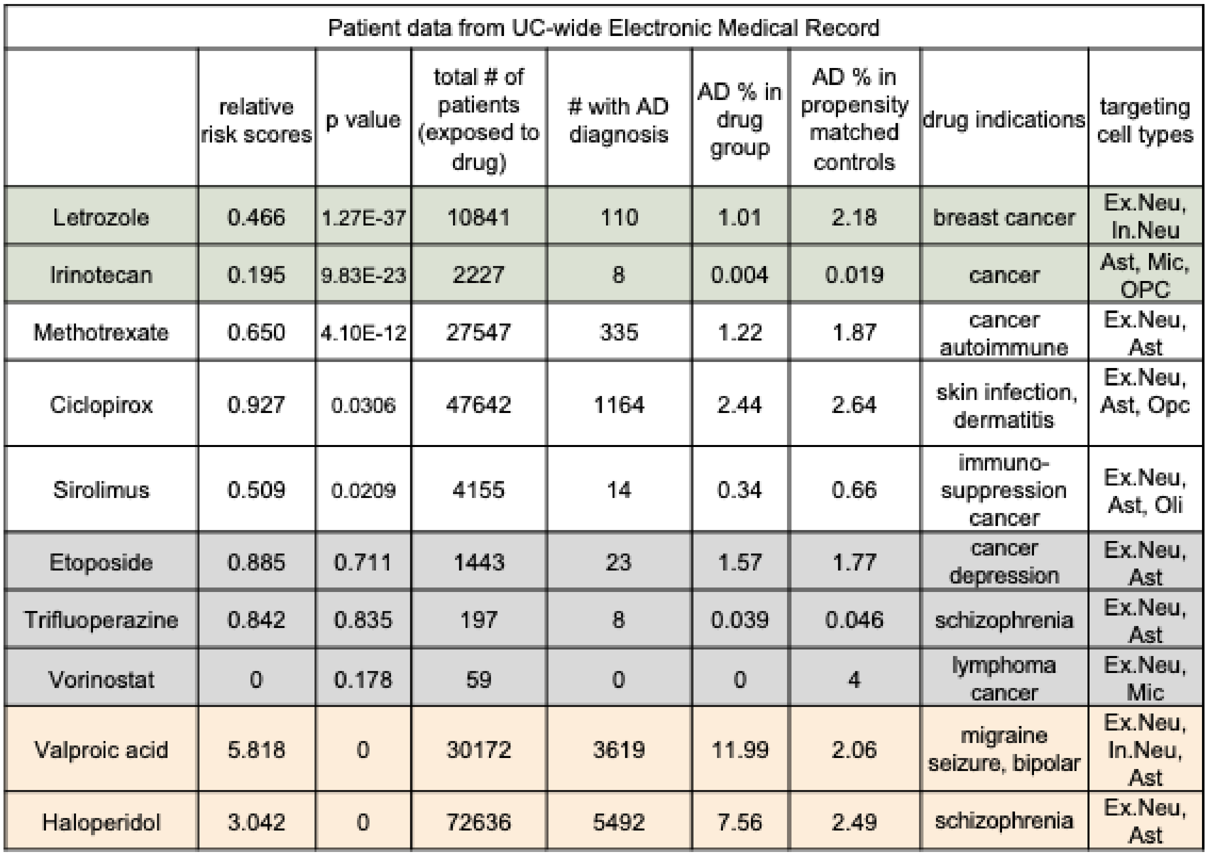
Table 2.5.1: Summary of cohort, sample, and feature sizes for individual and integrated datasets

Study	Cohort size	Total cell number	Number of features
Mathys et al.	24 AD, 24 ctl	70,634	17926
Zhou et al.	11AD, 11 ctl	72,489	33694
Lau et al.	12 AD, 9 ctl	165,959	33538
Integrated dataset (standardized and filtered case & control)	37 AD, 29 ctl	137,065	29120

Table 2.5.2: Case-control standardization reference

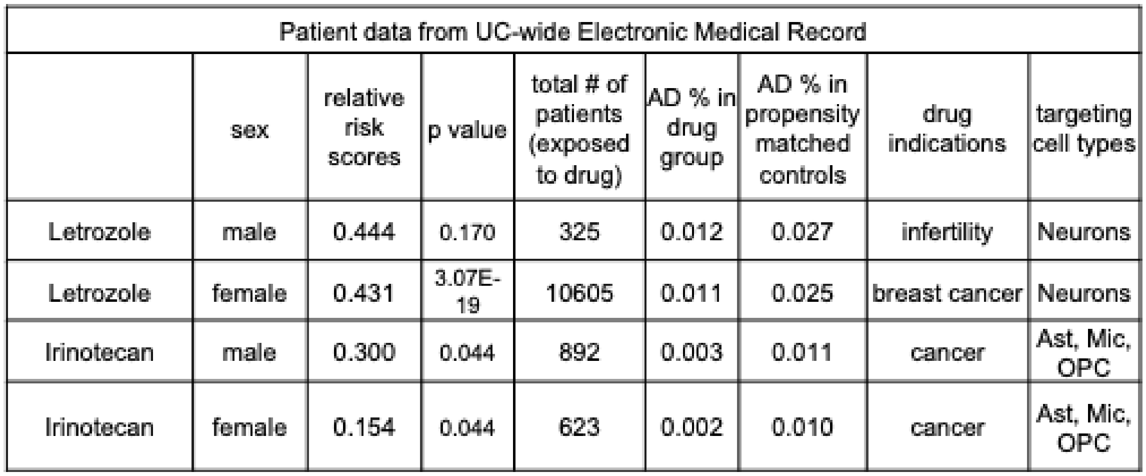
CERAD		Braak	
score	Coding	score	Severity
1	Definite	V & VI	Dementia
2	Probable	III & IV	Mild symptoms
3	Possible	I & II	Asymptomatic
4	No AD		

Table 2.5.3: AD outcome measures after drug exposure



Patient data from UC-wide Electronic Medical Record								
	relative risk scores	p value	total # of patients (exposed to drug)	# with AD diagnosis	AD % in drug group	AD % in propensity matched controls	drug indications	targeting cell types
Letrozole	0.466	1.27E-37	10841	110	1.01	2.18	breast cancer	Ex.Neu, In.Neu
Irinotecan	0.195	9.83E-23	2227	8	0.004	0.019	cancer	Ast, Mic, OPC
Methotrexate	0.650	4.10E-12	27547	335	1.22	1.87	cancer autoimmune	Ex.Neu, Ast
Ciclopirox	0.927	0.0306	47642	1164	2.44	2.64	skin infection, dermatitis	Ex.Neu, Ast, Opc
Sirolimus	0.509	0.0209	4155	14	0.34	0.66	immuno-suppression cancer	Ex.Neu, Ast, Oli
Etoposide	0.885	0.711	1443	23	1.57	1.77	cancer depression	Ex.Neu, Ast
Trifluoperazine	0.842	0.835	197	8	0.039	0.046	schizophrenia	Ex.Neu, Ast
Vorinostat	0	0.178	59	0	0	4	lymphoma cancer	Ex.Neu, Mic
Valproic acid	5.818	0	30172	3619	11.99	2.06	migraine seizure, bipolar	Ex.Neu, In.Neu, Ast
Haloperidol	3.042	0	72636	5492	7.56	2.49	schizophrenia	Ex.Neu, Ast

Table 2.5.4: Sex-stratified AD outcome measures after exposure of letrozole or irinotecan



Patient data from UC-wide Electronic Medical Record								
	sex	relative risk scores	p value	total # of patients (exposed to drug)	AD % in drug group	AD % in propensity matched controls	drug indications	targeting cell types
Letrozole	male	0.444	0.170	325	0.012	0.027	infertility	Neurons
Letrozole	female	0.431	3.07E-19	10605	0.011	0.025	breast cancer	Neurons
Irinotecan	male	0.300	0.044	892	0.003	0.011	cancer	Ast, Mic, OPC
Irinotecan	female	0.154	0.044	623	0.002	0.010	cancer	Ast, Mic, OPC

CHAPTER 3: SYSTEMATIC CHARACTERIZATION OF MOLECULAR INFLUENCE BY MAJOR RISK FACTORS IN MOUSE MODELS OF ALZHEIMER'S DISEASE

3.1 Introduction

Over 55 million people in the world have Alzheimer's Disease (AD) or related dementia and it is estimated that there are 416 million people living with prodromal and preclinical AD, and AD dementia[11]. Worldwide AD prevalence is higher in women: in the US, two-thirds of AD patients are women[74]. Besides sex, other non-modifiable risk factors for developing AD include age and genetics[75]. More specifically, more than 50% of patients with AD carry the *ε4* allele of the *APOE* gene, the most common genetic risk factor for late-onset AD[76]. *APOE4* carriers exhibit exacerbated AB plaques as well as neurofibrillary tangles, the two pathological hallmarks of AD[77]. Enabled by single-cell omics, mounting evidence and research suggests that *APOE4* has a multi-cellular effect on AD. For instance, an *APOE4*-stratified human single-nucleus transcriptomics study identified cell-type-specific enriched pathways in cells expressing APOE4[76].

Additionally, the disease progression and pathophysiology of AD differs by biological sex, with women patients having a much higher risk of developing AD compared to men, and as women age, their AD prevalence increases in comparison to men[74]. However, male patients' disease progression is more aggressive, and female AD patients live longer, highlighting the differing effect that age has in each sex. Further differences are apparent in the *APOE4* vulnerability, as women are at higher risk of developing AD when carrying the *APOE4* allele at earlier ages[77]. The sex and *APOE4* effect is also represented in AD's pathology: higher levels of total and hyperphosphorylated tau are more strongly associated with female *APOE4* carriers than male carriers[78]. Female *APOE4* carriers with mild cognitive impairment (MCI), the clinical stage prior to dementia, have also demonstrated greater hippocampal atrophy than male carriers with MCI[79].

At a cellular level, a single-cell transcriptomics' analysis on AD patients' brains revealed sex differences predominantly in glial cells and in neuroinflammatory pathways, beginning to elucidate potential mechanisms behind the sex dimorphism in AD[80]. This has also been corroborated by an *App* mouse model that showed that female microglia progressed into activated response microglia at earlier ages in comparison to male mice[81]. Mouse studies have also shown female-specific microglial response to *APOE4*: female microglia from mice carrying the human *APOE4* and overexpressing amyloid-beta, showed decreased Aβ plaque coverage, in comparison to their male and *APOE3* counterparts, using imaging[82]. Nevertheless, most sex-stratified studies have focused on only studying microglia, and we have yet to understand how all cell types are affected by sex and *APOE4*.

Despite wide recognition of age, *APOE4* and sex as primary risk factors of AD, the molecular underpinnings of how these risk factors affect the pathophysiology of AD are poorly understood. This chapter aims to assess the complex interplay between age, sex and *APOE4* genotype in regulating cell type-specific transcriptional profiles in a mouse model of AD. We perform single-nucleus transcriptomics on a sex-balanced cohort of homozygous human *APOE4* knock-in and human *APOE3* knock-in mice.

3.2 Results

To systematically investigate the specific and interactive effects of the three major risk factors associated with AD, we generated a comprehensive snRNA-seq dataset comprising hippocampal samples from twelve experimental subgroups (with n = 3-4 mice per condition) with varying conditions in age (at 6, 12 or 18 months of age), sex (male or female), and human homozygous *APOE* genotype (*APOE* 3/3 or *APOE* 4/4). While other non-neuronal cell types were detected, our primary focus was directed towards six disease-relevant brain cell types including excitatory neurons (Ex.Neu), inhibitory neurons (In.Neu), astrocytes

(Ast), microglia (Mic), oligodendrocytes (Oli), and oligodendrocyte precursor cells (OPCs). Acknowledging the selective vulnerabilities among various neuronal subtypes, we further subcategorized neuronal cells into CA1 pyramids, CA3 pyramids, dentate gyrus granule cells, subiculum neurons, and interneurons. Our analysis encompassed a risk-factor stratified approach, incorporating assessments of cell abundance, intercellular communications, differential gene expression, and pathway enrichment across the twelve distinct experimental conditions (**Figure 3.4.1**).

After standard processing and quality control, we obtained a filtered dataset containing 27,153 gene features expressed across 536,498 nuclei for further analysis. Through graph-based clustering and visualization using Uniform Manifold Visualization (UMAP), we identified 35 distinct cell clusters (**Figure 3.4.1**). The annotation of these clusters into either neuronal subtypes or non-neuronal cell type were achieved by comparing their expression patterns to known marker gene sets (**Figure 3.4.2**). The clusters were clearly separated according to the six major cell types (**Figure 3.4.1**) and exhibited an even distribution across different ages, sexes and *APOE* genotypes (**Figure 3.4.2**). UMAP labeling based on the twelve experimental conditions unveiled a moderately balanced distribution of nuclei across clusters for all conditions (**Figure 3.4.1**). As predicted, the majority of the human *APOE* expression resided within the astrocyte and microglia clusters, while negligible mouse *Apoe* expression was detected (**Figure 3.4.2**).

The three-way ANOVA test was conducted to assess the effects of age, sex, and *APOE* genotype, as well as their interactions, on significant differences observed in cell type abundance across the twelve experimental conditions. Cell type abundance was quantified as a fraction of the respective cell type within each biological samples. Interestingly, sex emerged as the most influential and only statistically significant factor contributing to the variance in astrocyte and inhibitory neuron abundances (**Figure 3.4.3**), with higher fractions observed in females for both cell types (**Figure 3.4.3**). While the interactive effect of age and *APOE* genotype was the second most influential factor, suggests a potential disparity in the impact of aging on astrocyte abundance across different *APOE* genotypes, this interactive effect did not achieve statistical significance (p-adj = 0.073). Furthermore, *APOE* genotype, along with the interaction between age and *APOE* genotype, emerged as the primary and statistically significant factors driving the observed variation in OPC proportions. Specifically, *APOE 4/4* carriers displayed higher OPC fractions at a younger age, with this difference diminishing as they aged. Regarding microglia, both sex and age emerged as the primary explanatory factors for variance. However, neither factor achieved statistical significance (**Figure 3.4.4**), possibly due to the limited sample size (n = 3-4).

Furthermore, we assessed the influence of individual risk factors and their interactive effects on cell abundance within distinct neuronal subpopulations. For each biological sample, the fraction of cells was quantified as cell counts within a specific cluster and normalized by the total cell count for that sample. Once more, we consistently observed sex to be a significant variable influencing factor for cell abundance among neuronal subpopulations. Specifically, higher fractions of CA1 pyramidal cells (cluster 24) were observed in females, while the opposite trend was exhibited in the dentate gyrus (cluster 3) and CA3 pyramids subpopulations (cluster 24 and 32) (**Figure 3.4.3**).

To further investigate cell type differences, we performed cell-cell communication (CCC) analysis and identified age-dependent sex differences in *APOE4* versus *APOE3* mice. When comparing overall CCC across cell types across aging, we observed decreased cell communication interaction strength between excitatory neurons and inhibitory neurons in *APOE4* male mice with age (**Figure 3.4.5**). In comparison, in female mice, the neuron-to-neuron communication does not change with time and remains higher in *APOE4* vs *APOE3*. Astrocytes also show increased CCC signaling strength to excitatory neurons in *APOE4* female mice in comparison to *APOE3*: In contrast, in males, astrocyte signaling decreases in adult and old mice in *APOE4* versus *APOE3* mice. Lastly, neuron-to-oligodendrocyte CCC strength increases in *APOE4* male mice as they age, a trend not observed in female mice.

When comparing female vs male mice (**Figure 3.4.5**), we also observed an overall decrease in female *APOE3* mice, and the opposite trend in female *APOE4* mice, as mice age, especially to and from excitatory and inhibitory neurons. Most cell-cell communication in young mice was increased in females regardless of *APOE* genotype, but as the mice aged, the cell-cell interaction strength decreased in females in *APOE3* mice only. These results suggest that female sex and *APOE4* genotype are key drivers of increased CCC across cell types, especially in neurons and astrocytes, and that age exacerbates the differences in sex and genotype.

To assess the age effect, we compared CCC between old (18 months) and adult (12 months) mice (**Figure 3.4.5**). It is apparent that CCC among neurons decreases with age in female *APOE3* mice and in both male and female *APOE4* mice, suggesting that female sex and age is enough to decrease CCC, whereas in male mice, *APOE4* is required to decrease CCC.

We then identified significant CCC pathways, defined as ligand-receptor pairs, stratified by cell type, sex, and age (**Figure 3.4.5**). NRG and NRXN pathways decrease in adult female mice in *APOE4* vs *APOE3*, whereas in males, these pathways increase in *APOE4*. Differences are especially apparent in astrocytes, neurons, oligodendrocyte precursors, and oligodendrocytes. The NRXN pathway is comprised of communication between Nrxn1 and Nrxn2 to Nlgn1 receptor, involved in synaptic contact and glutamatergic neurotransmission, and has been shown to be decreased in CSF and brain samples of AD patients[83]. The NRG pathway represents ligands Nrg1, Nrg2 and Nrg3, to ligand Erbb4, a receptor tyrosine kinase that has been implicated in GABAergic transmission and that is increased in human AD brains and *APP/PS1* transgenic mice[84].

To comprehensively understand the complex interplay of age, sex, and *APOE* genotype influencing transcriptomic processes in disease development, we aim to systematically compare 12 distinct conditions, each characterized by a unique combination of these three risk factors. Although we originally intended to conduct every possible set of parallel comparisons between two conditions differing in a single risk factor, stratified by cell types, this approach generates a substantial amount of data that becomes challenging to interpret.

While advanced age and the female sex exert significant influence over disease etiology and individual susceptibilities, the distinctions within the *APOE* genotype assume a more direct and disease-specific relevance in the context of AD. Therefore, we prioritized direct comparison between *APOE* genotypes. We utilized samples harboring the risk-associated human *APOE4* alleles to represent the disease state, while *APOE3* carriers served as proxies for control state. With this approach, we investigated the impacts of aging and sex on gene expression changes resulting from *APOE* genotype differences.

To accomplish this, we identified cell-type-specific differentially expressed genes (DEGs) by comparing *APOE4/4* samples to *APOE3/3* samples within each distinct age and sex conditions. The profiling of *APOE4/4* to *APOE3/3* expression changes exhibited significant variation across ages and between sexes (**Figure 3.4.6**). In general, we observed a higher transcriptomic difference between *APOE4/4* and *APOE3/3* males compared to females, as evidenced by the increased DEG counts across all ages and in most cell types. However, exceptions were noted for inhibitory neurons and microglia at 18 months old, where more DEGs were observed in females. Additionally, we observed an increased number of downregulated genes in neurons during old age (18 months), except in female excitatory neurons. The number of DEGs in glial cells tended to peak at middle age (12 months) and decrease at old age, except for microglia, which peaked at old age.

We performed a cross-cell type comparison of DEGs at 18 months old and found little overlaps across cell types within male and female. In male, more DEG overlaps across cell types were regulated in the same direction in *APOE4* samples versus *APOE3* samples. In females, the biggest DEG intersects contain four genes that were up regulated in astrocyte and down regulated in inhibitory neurons (**Figure 3.4.6**). Additionally, we investigated pathway enrichment by APOE4 genotype in male and females at 18 months old and visualized the results as networks of biological cluster. Majority of the *APOE4* versus *APOE3* pathway enrichment in males were downregulated, except cytoplasmic translation, regulation of cellular organelle, and nitric oxide activities were upregulated in oligodendrocytes. In contrast, more biological process was enriched by the upregulated *APOE4* versus *APOE3* DEGs in females, such as enrichment of glutamate signaling in astrocytes, adhesion cell factors in inhibitory neurons, and morphogenesis in microglia. These findings suggest *APOE4* genotype leads to highly divergent cell-type-specific transcriptomic response in male and in female, specifically, higher activation of immune cells and pathways related to hyperexcitability were observed in females (**Figure 3.4.6**).

Microglial cells exhibited distinctive characteristics when compared to other cell types within the context of the study. Notably, there was a limited number of DEGs observed between *APOE4* and *APOE3* conditions during young and middle age, while a substantial increase in DEGs was evident during old age, accompanied by pronounced sexual dimorphism, as illustrated in **Figure 3.4.7**. Specifically, among aged microglia, the intersection of DEGs between females and males was quite minimal, with only three downregulated DEGs (Meg3, Nrg3os, and Gm20642) and one upregulated DEG (GM42418) being shared. Notably, an intriguing observation pertained to the identification of three DEGs that were shared between females and males during middle age but exhibited opposite patterns of regulation. These DEGs were Fkbp5, implicated in cellular stress response, as well as Zbtb16 and Ccnd3, both of which play pivotal roles in cell proliferation processes (**Figure 3.4.7**).

The count *APOE4* versus *APOE3* DEGs of Astrocyte peaked at 12 months old in both males and females. The pathway enrichments by these DEGs differ in the direction of regulation in male versus females, and were mostly related to cell adhesion, developmental process, cell junction assembly, and intracellular signal transduction. Minimal DEG overlaps were observed between females and males at 18 months old, except glutamate binding, opposite directions of gene regulations between sexes (**Figure 3.4.7**).

In inhibitory neurons, the *APOE4* versus *APOE3* differential gene expression patterns showed a strong age-dependency, with 283 DEGs upregulated in males at 6 months old and only 12 DEGs at 18 months old (**Figure 3.4.8**) Additionally, large number of DEGs in 6 months old males overlapped but regulated in opposite directions with 18 months old males as well as 6- and 18-months old females (**Figure 3.4.8**).

Glumatergic signaling pathway and synapse transmission were upregulated in 18 months females while glutamate binding and calcium channel activity were downregulated in 18 months males (**Figure 3.4.8**).

Differential gene expressions in OPC due to APOE4 genotype was the strongest at 12 months old in both males and females. Overall, DEGs in males outnumbered DEGs in females at all age groups. (**Figure 3.4.9**) We observed significant pathway enrichments by the downregulated genes associated with cell-cell signaling in male OPC (**Figure 3.4.9**), consistent with the cell-cell-communication analysis in **Figure 3.4.5**.

3.3 Methods

cDNA Library Preparation and Sequencing

Single-nuclei preparation for 10x loading. One frozen mouse hippocampus was placed into a pre-chilled 2 mL Dounce with 1 mL of cold 1X Homogenization Buffer (1X HB) (250 mM Sucrose, 25 mM KCL, 5 mM $MgCl_2$, 20 mM Tricine-KOH pH7.8, 1 mM DTT, 0.5 mM Sermidine, 0.15 mM Sermine, 0.3% NP40, 0.2 units/µL RNase inhibitor, ~0.07 tabs/ml cOmplete Protease inhibitor). Dounce with "A" loose pestle (~10 strokes) and then with "B" tight pestle (~10 strokes). The homogenate was filtered using a 70 µM Flowmi strainer (Bel-Art) and transferred to a pre-chilled 2 mL LoBind tube (Fischer Scientific). Nuclei were pelleted by spinning for 5 min at 4°C at 350 RCF. The supernatant was removed, and the nuclei were resuspended in 400 µL 1X HB. Next, 400 µL of 50% Iodixanol solution was added to the nuclei and then slowly layered with 600 µL of 30% Iodixanol solution under the 25% mixture, then layered with 600 µL of 40% Iodixanol solution under the 30% mixture. The nuclei were then spun for 20 min at 4°C at 3,000RCF in a pre-chilled swinging bucket centrifuge. 200 µL of the nuclei band at the 30%-40% interface was collected and transferred to a fresh tube. Then, 800 µL of 2.5% BSA in PBS plus 0.2 units/µL of RNase inhibitor was added to the nuclei and then were spun for 10 min at 500 RCF at 4C. The nuclei were resuspended with 2% BSA in PBS plus 0.2 units/µL RNase inhibitor to reach at least 500 nuclei/µL. The nuclei were then filtered with a 40 µM Flowmi cell strainer. The nuclei were counted and then ~13,000 nuclei per sample were loaded onto 10x Genomics Next GEM chip M. The snRNA-seq libraries were prepared using the Chromium Next GEM Single Cell 3′ HT kit v3.1 (10x Genomics) according to the manufacturer's instructions. Libraries were sequenced on an Illumina NovaSeq 6000 sequencer at the UCSF CAT sequencing core.

Sequence Alignment, Filtering, and Counting

The demultiplexed fastq files were processed following the procedure previously described by Zalocusky et al. In summary, the fastq files were aligned to a custom reference genome, built from mm10-1.2.0 which includes introns, using the *cellranger count* function (version 4.0.0) with default parameters, as detailed in the Cell Ranger documentation. Subsequently, a single UMI count file per animal/sample was generated by the *cellranger count* function. Individual UMI count files were then combined into a single count matrix using *merge* function in Seurat package v4.0.4. Metadata, including age, sex, and genotype information, were added to each cell.

Pre-processing and Quality Control

The count matrix was further processed with Seurat by first calculating percentage of mitochondria genes mapped per cell. The distribution of feature count, total mapped gene count, and percentage of mitochondria genes were visualized across biological samples as violin plots, and no obvious outlier was identified. We filtered the count matrix to only include cells with higher than 250 gene features, at least 500 gene counts, and mitochondria gene percentage lower than 10%. Potential misaligned or ambiguous gene features that expressed in fewer than 10 cells were also removed. These quality assurance steps resulted in a final Seurat object containing 27,153 gene features expressed by 536,498 nuclei.

Normalization, Dimensional Reduction, and Clustering

Count normalization and dimensionality reduction was conducted following standard procedure in Seurat package. In brief, we performed normalization and variance stabilization with an updated version of *sctransform*, v2, and principal component analysis (PCA) with *RunPCA* (npcs = 30). Dimensional reduction through Uniform Manifold Approximation and Projection (UMAP) was performed with the *RunUMAP* function and considering the top 15 dimensions selected from the corresponding PCA.

The clustering was based on the first 15 principal components (PCs) using the *FindNeighbors* function in Seurat. This function embeds cells in a K-nearest neighbor graph, considering the Euclidean distance in PCA space and refining the edge weights between any two cells based on the shared overlap in their local neighborhoods. The clustering was obtained using the *FindClusters* function, which employs modularity optimization techniques such as the Louvain algorithm, with a resolution parameter of 0.5 and resulted in a set of 35 distinct clusters.

Cell-Type Annotation

Major cell types, including astrocytes, microglia, oligodendrocytes, oligodendrocyte precursor cells, excitatory and inhibitory neurons, were classified using mouse brain cell markers in PanglaoDB, a publicly available marker gene database. Further subdivision of hippocampus cell types, such as CA1 and CA3 pyramidal cells, were queried against hippocampal cell-type-specific marker genes as published in hipposeq (https://hipposeq.janelia.org). Cell type identities per cluster were determined by applying Seurat's *AddModuleScore* function to sets of mouse brain marker genes. A module score for each cell type considered was calculated per cell. Each cell was assigned the corresponding cell type identity that generated the highest scores among scores for all cell types. If the highest and second highest scores of a cell were within 20% of the highest score, then the cell were deemed hybrids and excluded from further analysis. We assessed the validity of the assigned cell type identities by examining the homogeneity, distribution, and separation of cell types by clustering in UMAP plots. Minority cell types in a cluster, defined as cell type that accounts for less than 5% of the total counts for that cluster, were considered as potential hybrid cells and excluded from further analysis.

Cell Abundance Analysis

To determine whether the risk factors of interest- age, sex, and genotype- contribute to the variance in cell type abundance, we compared cell fractions per biological sample within each cluster across different sample groups. Briefly, the fraction calculation involved dividing the cell count of a biological sample in a specific cluster by the total cell counts for that sample across the entire dataset. Within each cluster, we grouped biological replicates (n = 3 or 4) and compared the distribution of these fractions across various sample groups using a three-way-ANOVA test. We employed the three-way-ANOVA test to investigate whether the independent variables (the three risk factors) or their interactive effects significantly contribute to the observed variation in outcomes, specifically, the difference in cell abundance. This analysis was conducted using the *aov* function within the Stats package (version 3.6.2) in R.

Cell Cycle Scoring and Annotation

To assess the variation in the distribution of cell phases within a specific cell type across different sample groups, we estimated the cell cycle phase of each cell. This was accomplished by calculating scores for each phase using the *CellCycleScoring* function in Seurat. This function predicts the classification of each cell into G1, G2M, or S phases based on lists of cell cycle markers obtained from Tirosh et al, as explained in the Seurat tutorial. Subsequently, cells were labeled with the cycle phase corresponding to the highest score. We computed the proportion of each cell cycle phase within a specific cell type and compared the proportions of each phase across sample groups. Finally, we applied three-way-ANOVA test to determine whether the risk factors under consideration- age, sex, or genotype - significantly contributed to the observed variance in cell phase abundance across sample groups.

Cell-Type-Specific Differential Expression Analysis

For each cell type, we performed a differential gene expression analysis comparing E44 cells to E33 cells and stratified the analysis by age and sex groups. This analysis was conducted using the *FindMarkers* function in the Seurat package. We set the test.use parameter to MAST, a two-part generalized linear model, as recommended by Mou et al. in a study comparing 9 DE methods for single-cell RNA sequencing analysis. MAST models gene expression rate using linear regression and expression level using Gaussian distribution. Differentially expressed genes (DEGs) were considered significant based on Bonferroni-corrected adjusted p-value cutoff of less than 0.05 and a log2 fold change (LFC) value greater than 0.1.

Pathway Analysis

We performed pathway enrichment analysis using a web tool g:Profiler, which conducts functional over-representation analysis by associating a given gene list with known functional information sources and identifies statistically significant enriched terms. For each cell-type-specific DEG list, we separated the upregulated and downregulated subsets and independently queried them with an adjusted p-value cutoff of 0.05 to identify significant pathway enrichments in both directions. We employed the g:SCS algorithm, a default method for multiple testing correction in pathway enrichment analysis, to account for dependencies among hierarchically related pathway terms. Our analysis encompassed terms from Gene Ontology cellular components, biological processes, and molecular functions, as well as Human Protein Atlas, Human Phenotype Ontology, KEGG, Reactome, and Wiki pathways.

Enriched Pathway Network Visualization

Following a previously established protocol, we conducted network pathway analysis on the pathway results derived from our cell-type-specific DEGs. In summary, the cell-type-specific upregulated and downregulated pathway were imported into the Cytoscape visualization application, EnrichmentMap, individually, and distinguished by color intensities. Upregulated pathways were labeled with dark colors, while downregulated pathways were labeled with light colors. Each node, representing a pathway, was filled with different color tones to indicate their cell type associations. We marked nodes green for excitatory neurons, blue for inhibitory neurons, yellow for astrocytes, red for microglia, purple for oligodendrocytes, and pink for oligodendrocyte precursor cells. Redundant and related pathways were consolidated into single biological themes using AutoAnnotate, a Cytoscape application. Pathway nodes are connected by edges if the two pathways share gene set overlaps. Labels for biological themed clusters were generated with WordCloud algorithm using the "Adjacent Words" option.

Cell-Cell Communication

Cell-cell communication (CCC) was calculated using the CellChat[85] method. Briefly, CellChat utilizes ligand, receptor, and cofactor expression from transcriptomic data to calculate a CCC probability. First, based on a CellChat-curated database of ligand-receptor interactions, differentially expressed signaling genes are used to calculate ensemble average expression of signaling genes. Communication probability is modeled using the law of mass action, and statistically significant communications are identified using a permutation test. We then evaluated the interaction across cell types (excitatory neurons, inhibitory neurons, oligodendrocytes, oligodendrocyte precursor cells, and astrocytes).

3.4 Figures

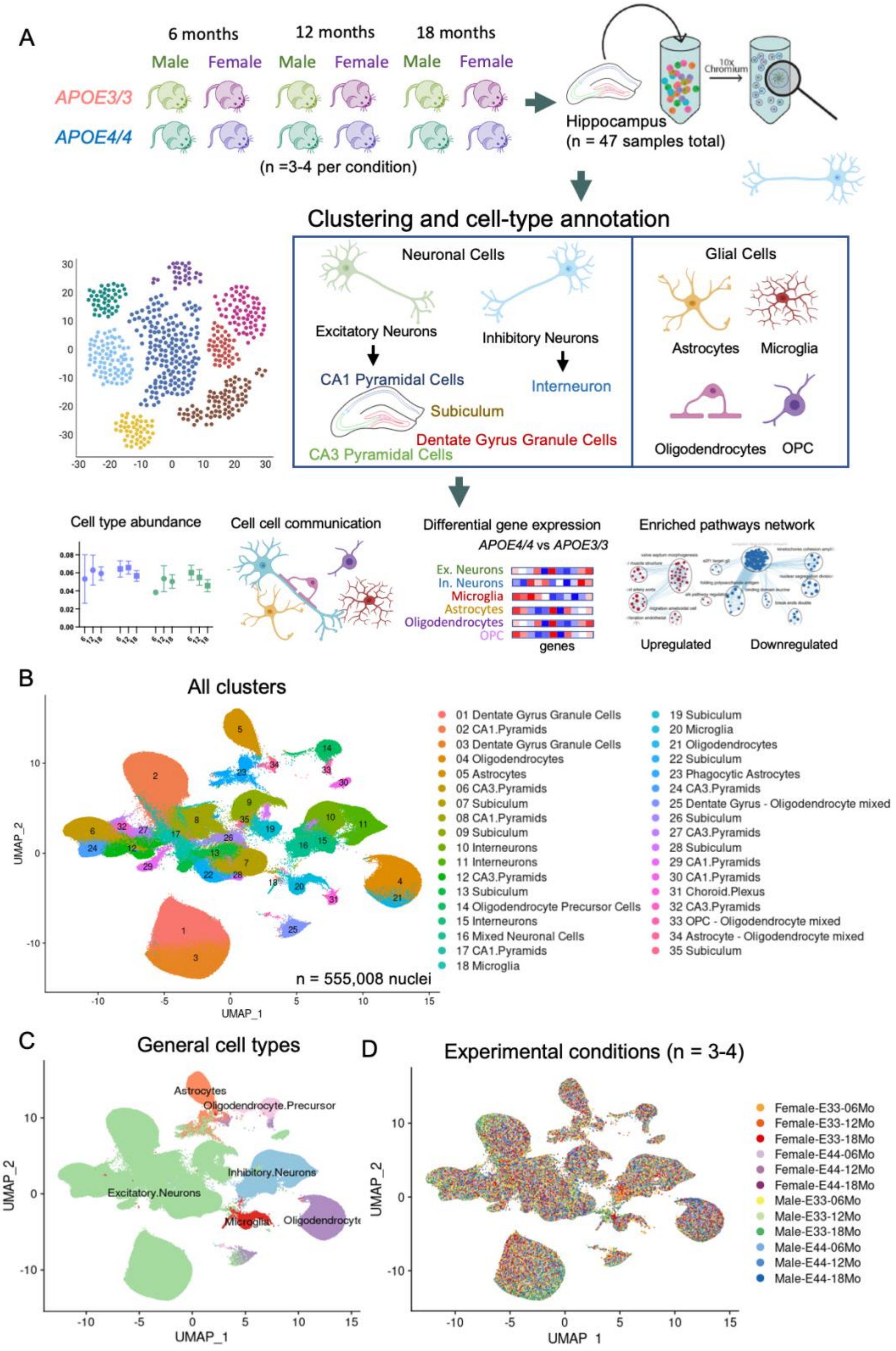


Figure 3.4.1 Study overview and dimensional reductions.

Figure 3.4.1 Study overview and dimensional reductions. A. Study overview including cohort design, RNA sequencing workflow of nuclei isolated from hippocampal samples, and computational analysis. B. UMAP clustering revealed 35 distinct cellular populations labeled by major cell types and neuronal sub-types. C. UMAP clustering annotated by prominent cell types that were used to stratify subsequent differential gene expression analysis. D. UMAP clustering annotated by the twelve distinct experimental conditions varying in three risk factors including age, sex, and genotype.

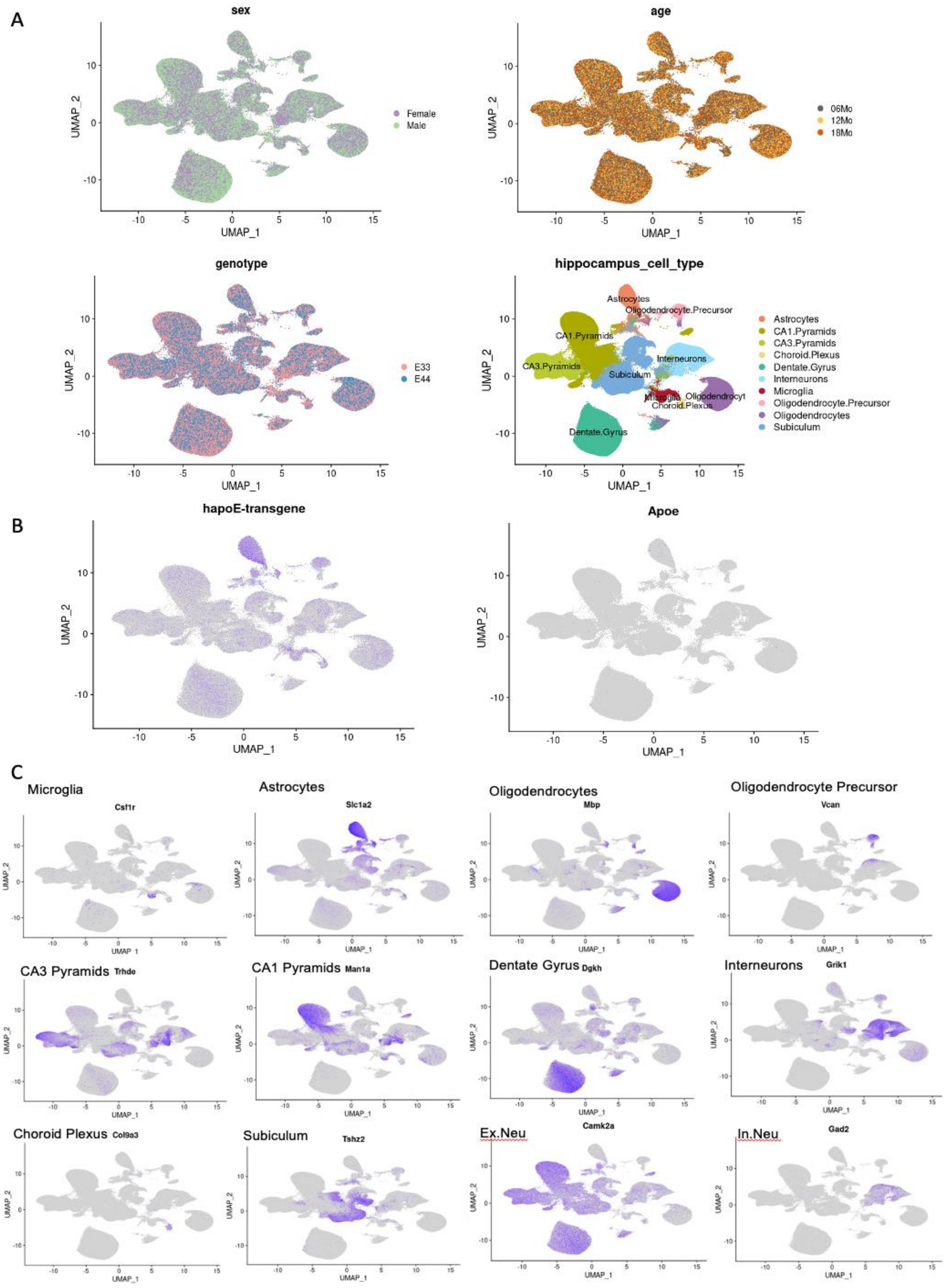


Figure 3.4.2 Extended dimensional reduction data.

Figure 3.4.2 Extended dimensional reduction data. A. UMAP clustering annotated by sexes, genotypes, ages and neuronal subpopulations separating regional locations in hippocampus. B. UMAP clustering with accentuated color to visualize expression pattern of human *APOE* gene and the mouse *Apoe* gene. C. Feature plots of expressions of marker genes for major cell types to visualize expression pattern and aid cell type identification.

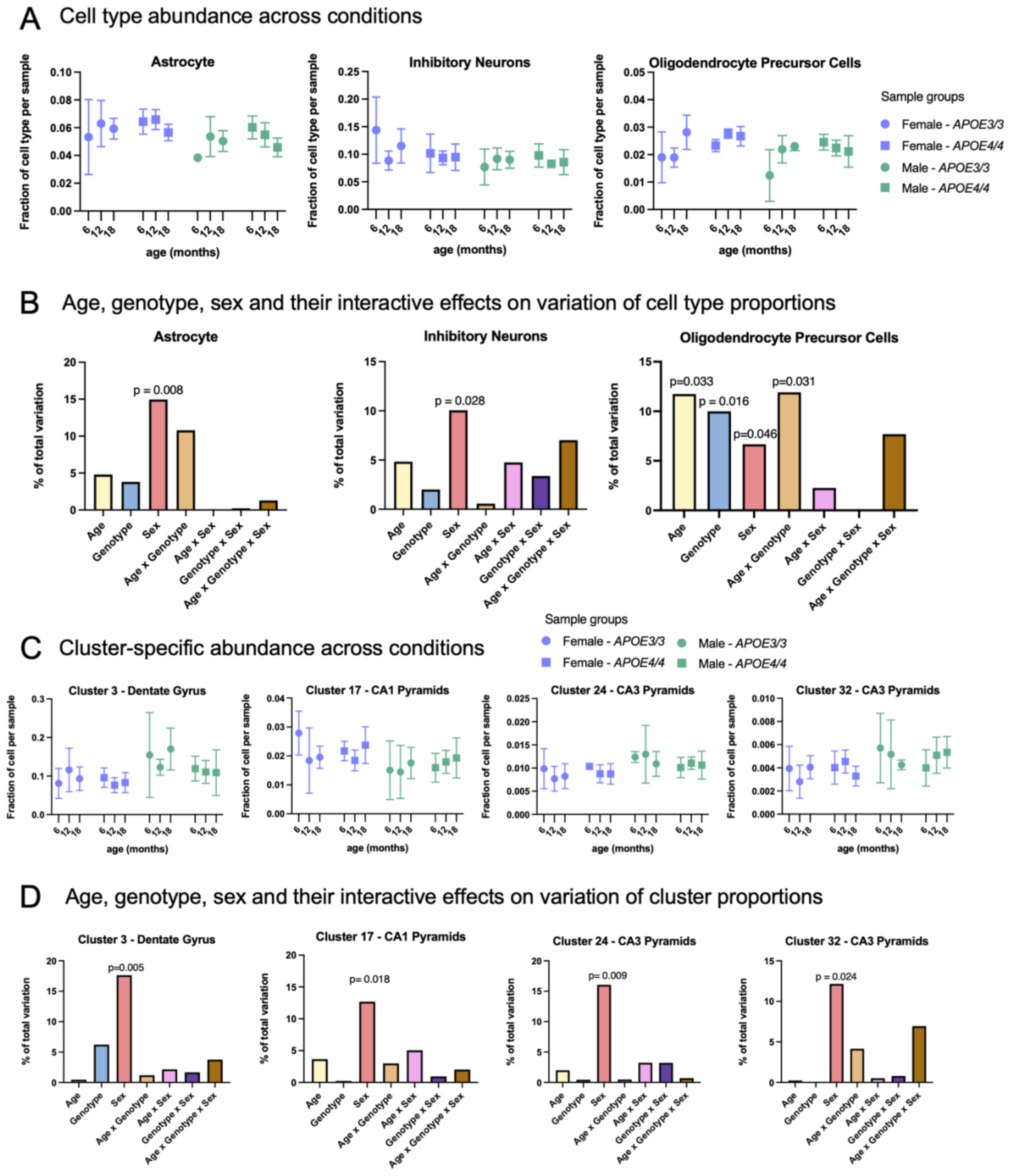


Figure 3.4.3 Cell abundance comparison across 12 conditions using three-way-ANOVA tests.

Figure 3.4.3 Cell abundance comparison across 12 conditions using three-way-ANOVA tests. The test considers individual risk factor effects and interactive effects on variance observed in cell fractions between groups. A. Cell type proportion per sample for astrocyte, inhibitory neurons, and OPC. B. Source of variance contributions to difference in cell type proportions across the 12 conditions. The heights of the bars represent amount of variance contributed by the risk factors or the interactive effects of two or more risk factors. P-value for significant variance contributor were indicated above the factor. C. Neuronal cell clusters with cell proportions significantly varied by sex across 12 conditions. D. Source of variance for the neuronal cell clusters with sex as a significant variance contributing factor. P-value was labeled above the risk factor-associated bars.

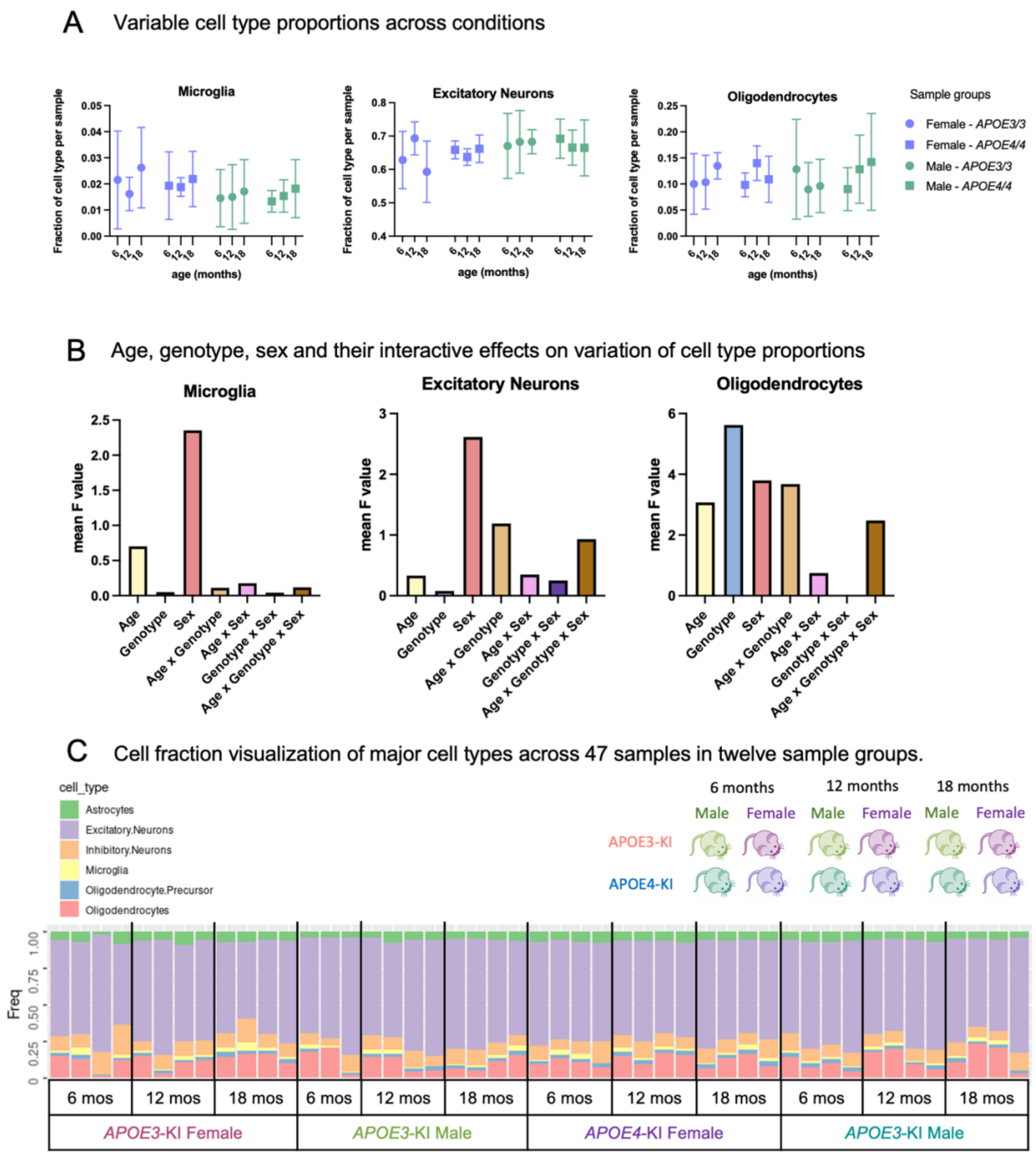


Figure 3.4.4 Cell type proportion extended data.

A. Cell type proportion per biological sample for microglia, excitatory neurons, and oligodendrocytes. B. Source of variance contributions to difference in cell type proportions across the 12 conditions. Bar heights represent amount of variance contribution by each individual risk factor and the interactions of two or two risk factors. No significance reached in the 3-way-ANOVA test. C. Cell type distribution within each sample and across 12 experimental conditions.

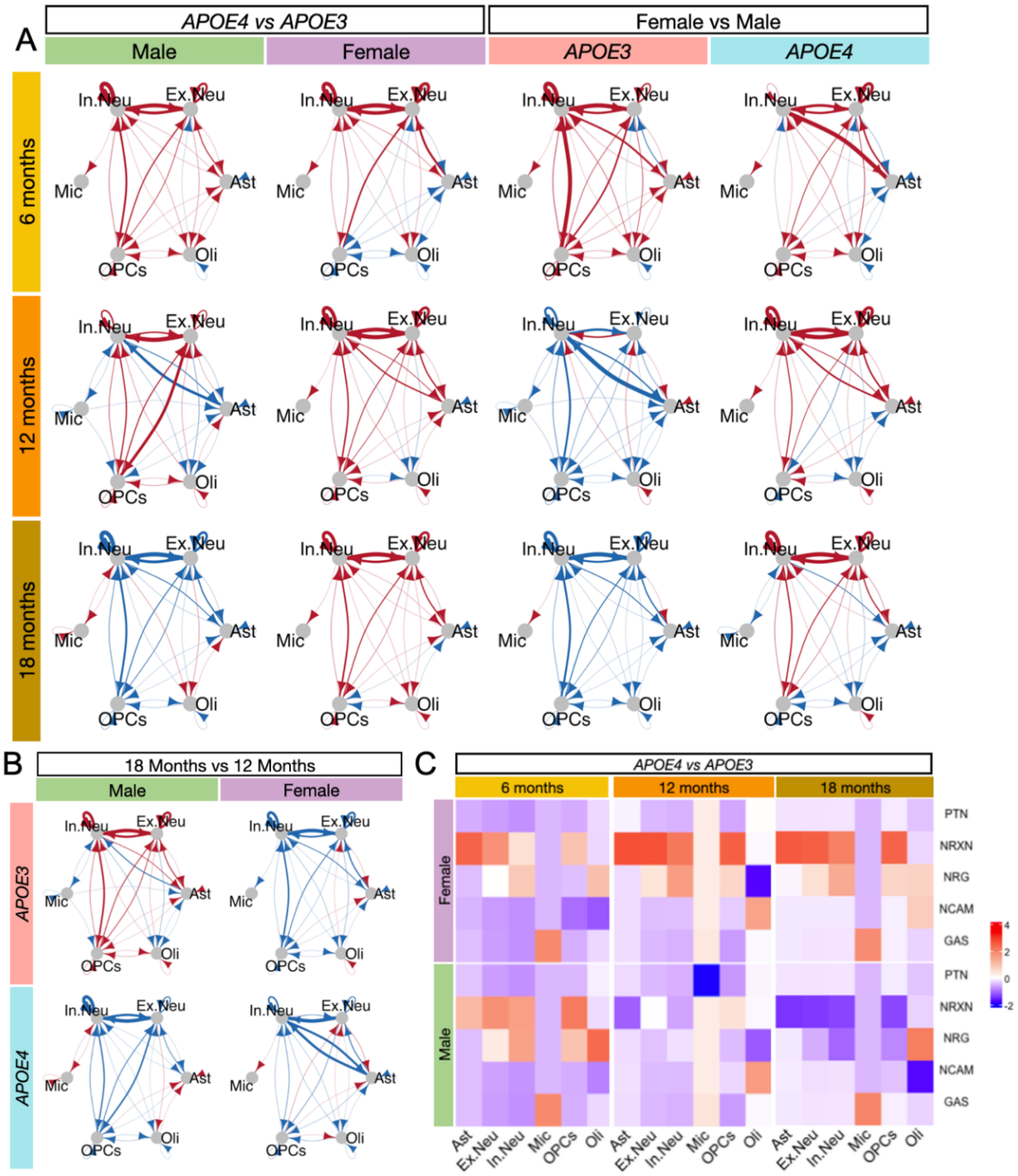
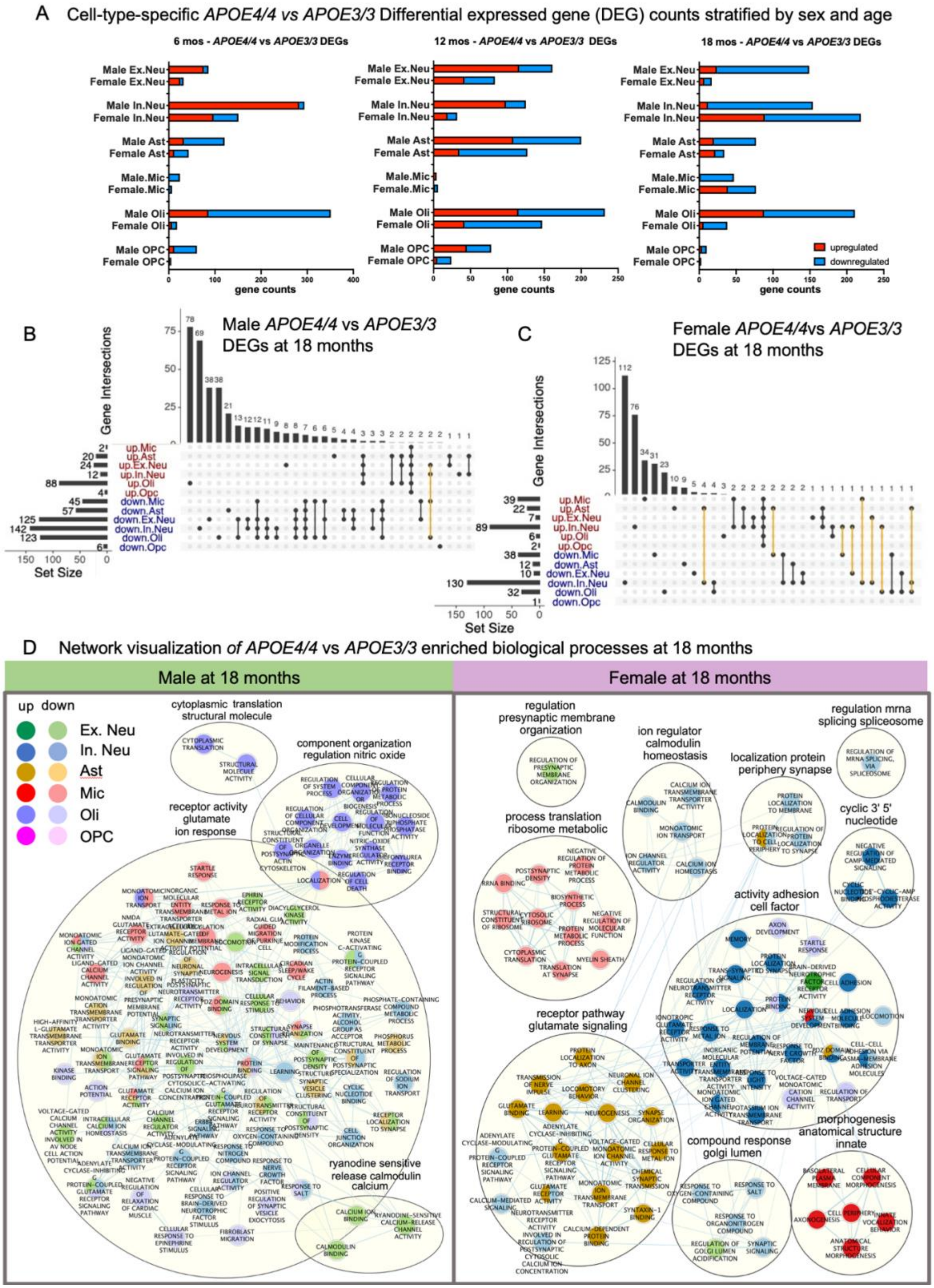


Figure 3.4.5 Cell-cell communication analysis.

Figure 3.4.5 Cell-cell communication analysis. A. Cell-cell communication (CCC) interaction strength evaluated by receptor-ligand pairs expression and represented as arrows connecting two cell types. Arrows represent the communication direction from sender to receiver cell type. Red connections in male/female columns indicate increased communication in *APOE4* vs *APOE3* and blue indicate decreased communications. Red connections in *APOE3/APOE4* columns indicate increased communication in female vs male. B. CC interaction strength difference when comparing old mice to adult mice. C. Heatmap displaying pathway-level CCC difference between in *APOE4* and *APOE3* across age and sex, stratified by cell type. Value represents the difference between *APOE4* and *APOE3* scaled CCC probability. Signaling pathways represent ligand-receptor pairs based on the KEGG Pathway database.



A Cell-type-specific *APOE4/4 vs APOE3/3* Differential expressed gene (DEG) counts stratified by sex and age

B Male *APOE4/4* vs *APOE3/3* DEGs at 18 months

C Female *APOE4/4*vs *APOE3/3* DEGs at 18 months

D Network visualization *of APOE4/4* vs *APOE3/3* enriched biological processes at 18 months

Figure 3.4.6 Differential gene expression analysis comparing *APOE4/4* vs *APOE3/3* samples.

Figure 3.4.6 Differential gene expression analysis comparing *APOE4/4* vs *APOE3/3* samples A. DEGs counts for six major cell types in male and females at 6, 12, and 18 months old. red bars represent upregulated genes, and blue bar represent downregulated genes. B,C. Upset plots showing DEG (*APOE4* vs *APOE3*) overlaps among six major cell types in males (B) and in females (C) at 18 months old. A red line separates DEG groups in the direction of regulation, either up and down-regulated. D. heatmap of a selection of intersecting DEGs across cell types in males, and differential expression pattern in females if differentially expressed. D. heatmap of all intersecting DEGs across cell types in females, and differential expression pattern in males. E. Enriched pathways networks for males and females at 18 mos. old with *APOE4* compared to *APOE3* and stratified by cell types. Network visualization of up or downregulated functional pathways in six major cell type. Pathways were grouped into biological themes and connected to other pathway clusters that share related mechanistic functions. Each node represents an enriched pathway and colored by the cell type or cell types associated. Lighter colors indicate downregulated pathways and darker colors indicate upregulated pathways.

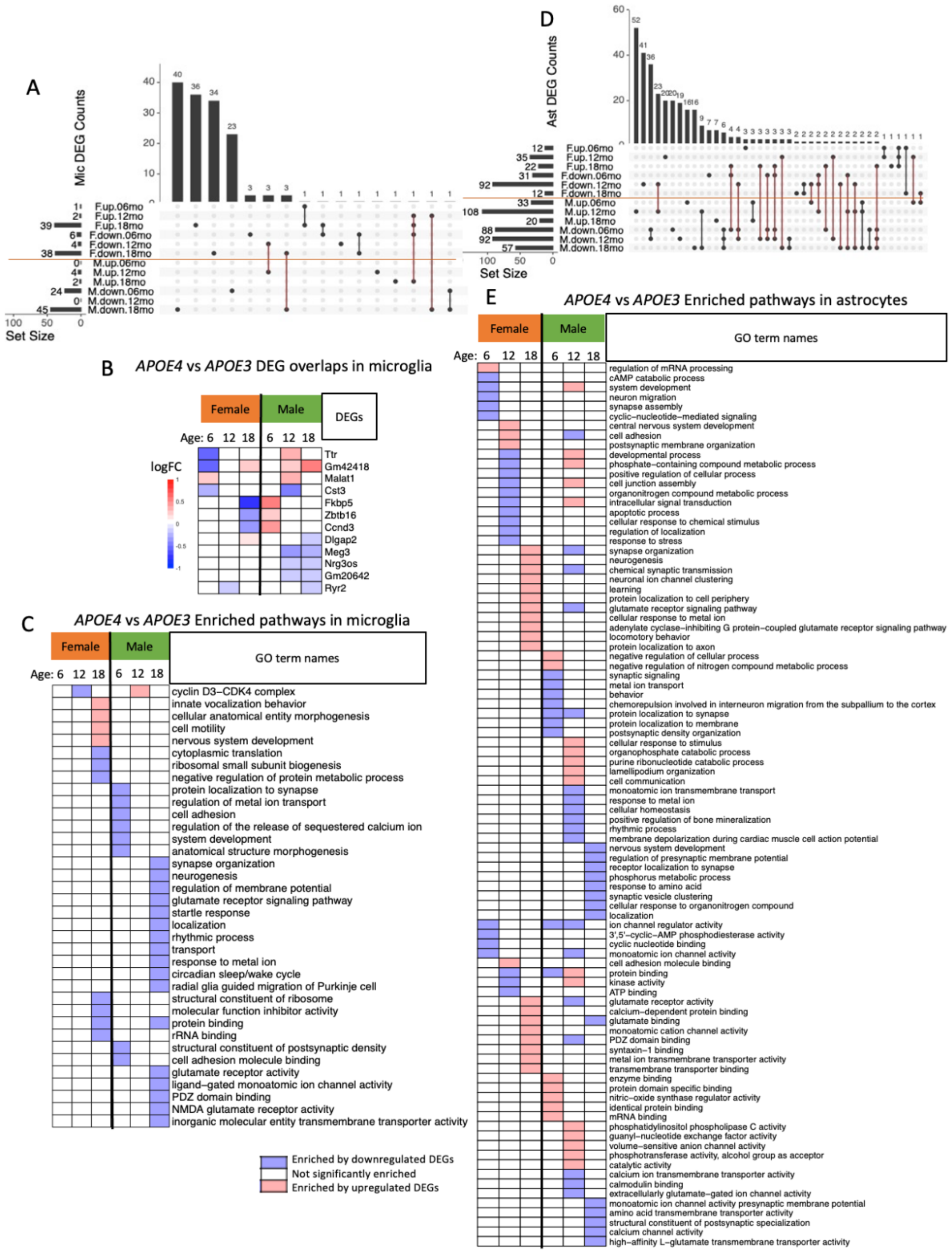


Figure 3.4.7 Microglia and astrocyte specific analysis.

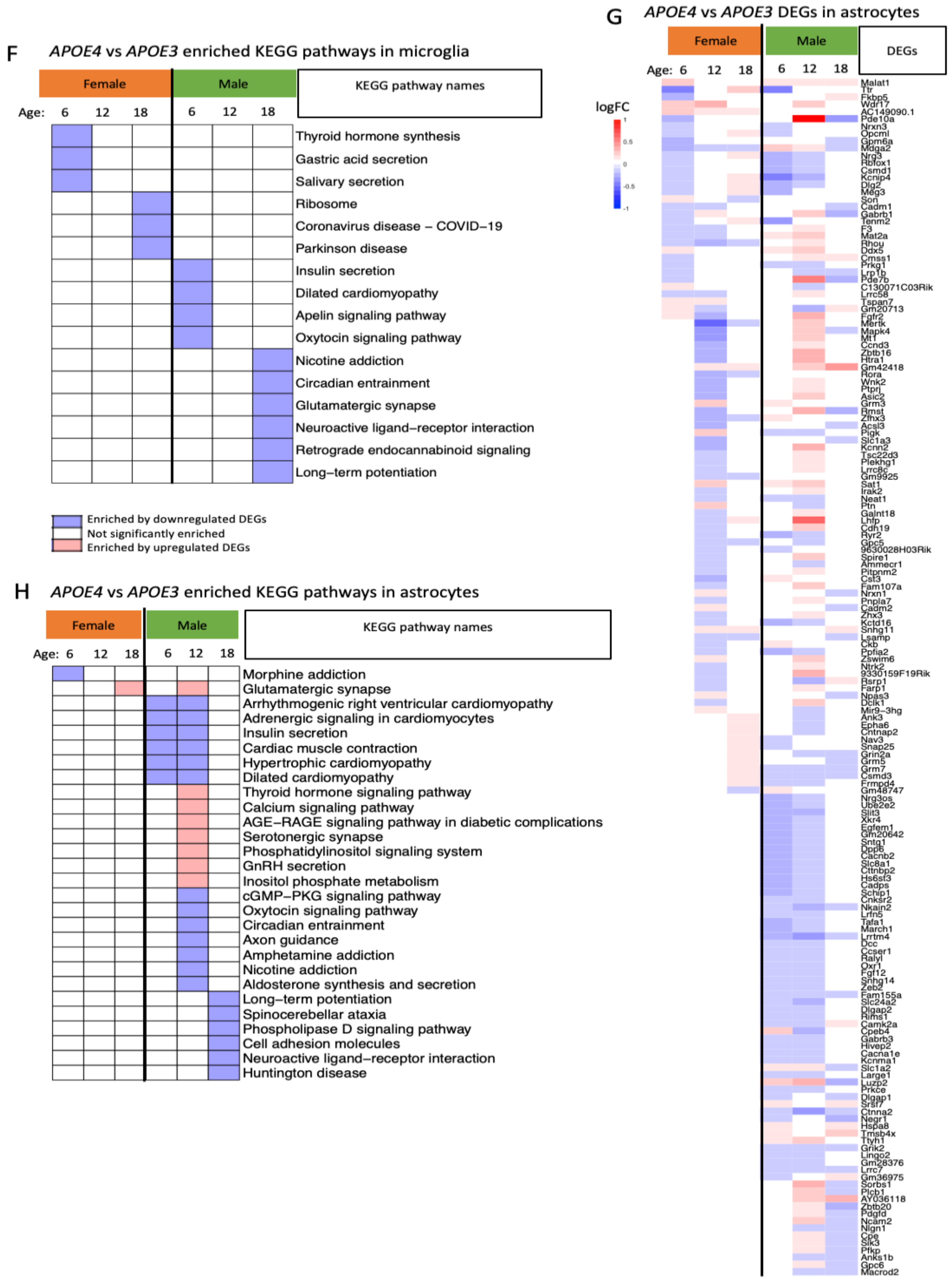


Figure 3.4.7 Microglia and astrocyte specific analysis.

Figure 3.4.7 Microglia and astrocyte specific analysis. A. Upset plots showing microglia DEG overlaps across sexes and ages, each condition further split by directions of gene regulations. red line crosses gene intersects between females and males. B. Heatmap of overlapped DEGs in males and females at each age. C. Heatmap showing *APOE4* vs *APOE3* enriched pathways in microglia stratified by age and sex. Colors indicate directions of regulation, with red labels for upregulated, blue labels for downregulated, and white labels non-significant in the associated groups. D. Upset plots showing astrocyte DEG overlaps across sexes and ages. E. Heatmap showing *APOE4* vs *APOE3* enriched pathways in astrocytes. F. Heatmap showing *APOE4* vs *APOE3* enriched KEGG pathways in microglia stratified by age and sex. Colors indicate directions of regulation, with red labels for upregulated, blue labels for downregulated, and white labels non-significant in the associated groups. G. Heatmap of overlapped astrocyte DEGs in males and females across age groups. H. Heatmap showing *APOE4* vs *APOE3* enriched KEGG pathways in astrocytes stratified by age and sex.

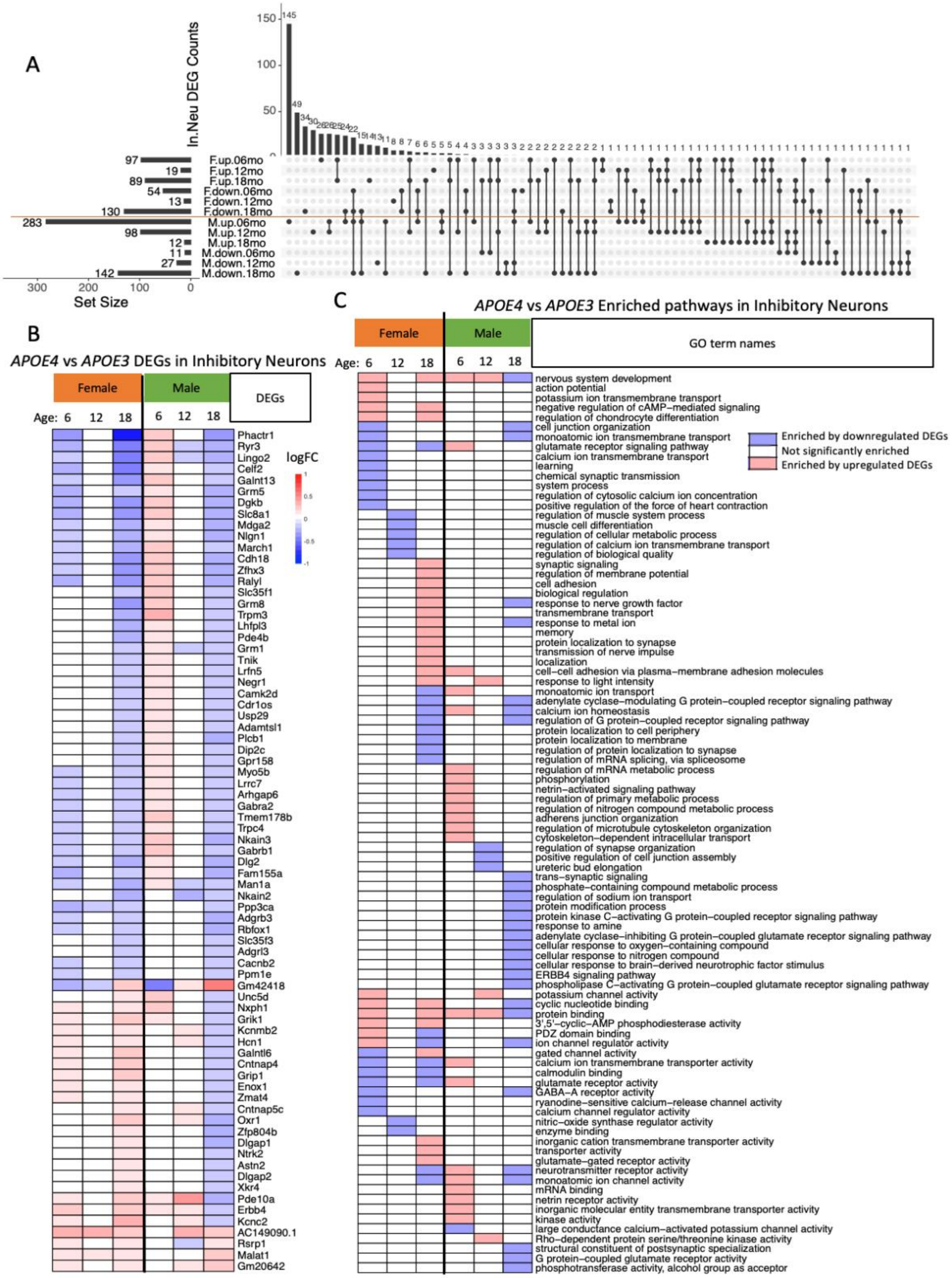


Figure 3.4.8 Inhibitory neurons specific analysis

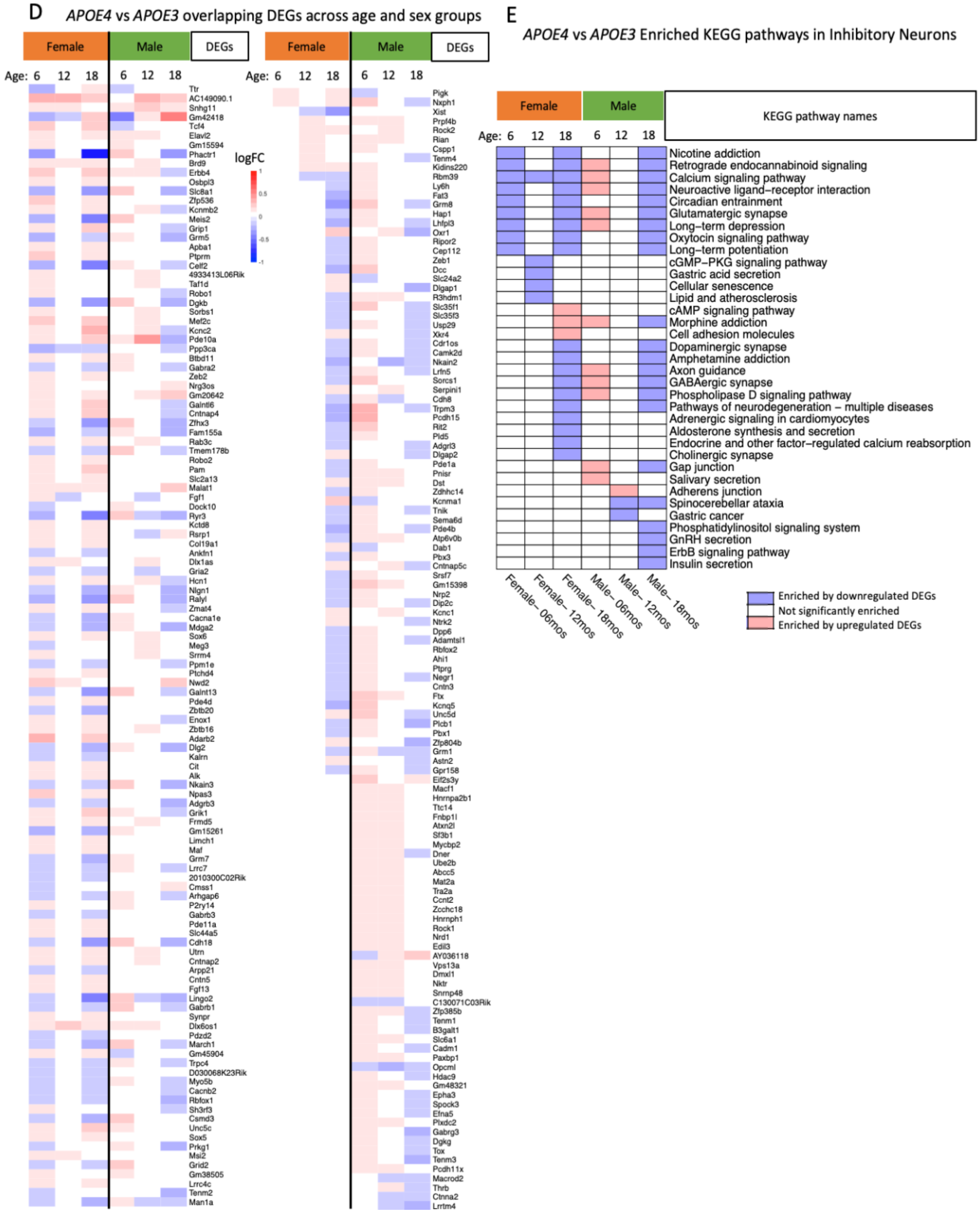
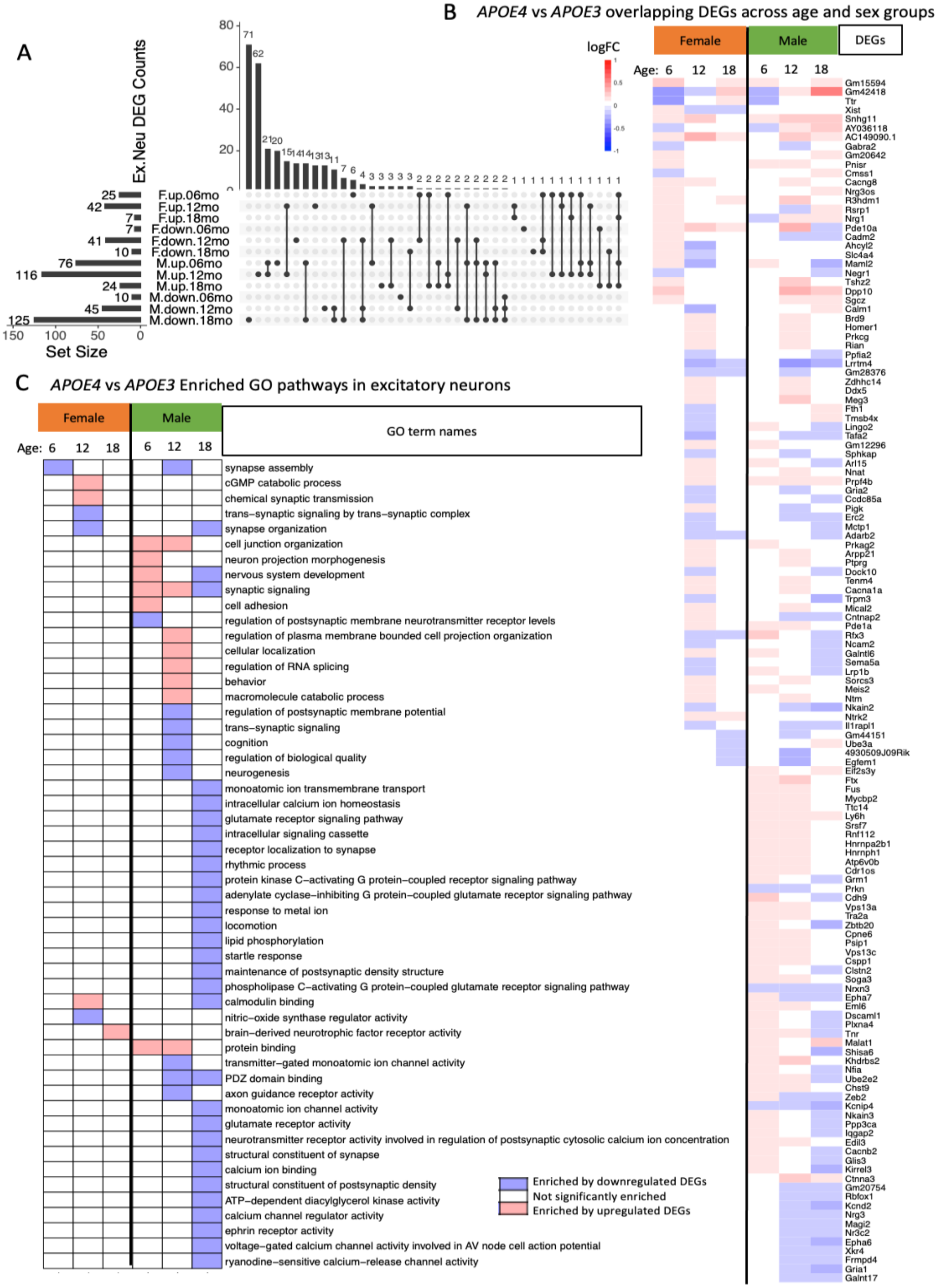


Figure 3.4.8 Inhibitory neurons specific analysis

Figure 3.4.8 Inhibitory neurons specific analysis A. Upset plots showing *APOE4* vs *APOE3* DEG overlaps in inhibitory neurons across sexes and ages, each condition further split by directions of gene regulations. red line crosses gene intersects between females and males. B. Heatmap showing expression of a selected subset of DEGs in inhibitory neurons that overlapped in males and females at 18 months of age. C. Heatmap showing *APOE4* vs *APOE3* enriched pathways in inhibitory neurons stratified by age and sex. Colors indicate directions of regulation, with red labels for upregulated, blue labels for downregulated, and white labels non-significant in the associated groups. D. Heatmap showing *APOE4* vs *APOE3* enriched KEGG pathways in Inhibitory neurons stratified by age and sex. Colors indicate directions of regulation, with red labels for upregulated, blue labels for downregulated, and white labels non-significant in the associated groups. E. Heatmap of overlapped DEGs of Inhibitory neurons in males and females across age groups.



C *APOE4* vs *APOE3* Enriched GO pathways in excitatory neurons

Female			Male			GO term names
Age: 6	12	18	6	12	18	
🟦				🟦		synapse assembly
	🟥					cGMP catabolic process
	🟥					chemical synaptic transmission
	🟦					trans–synaptic signaling by trans–synaptic complex
	🟦				🟦	synapse organization
			🟥	🟥		cell junction organization
			🟥			neuron projection morphogenesis
			🟥		🟦	nervous system development
			🟥	🟥	🟦	synaptic signaling
			🟥			cell adhesion
			🟦			regulation of postsynaptic membrane neurotransmitter receptor levels
				🟥		regulation of plasma membrane bounded cell projection organization
				🟥		cellular localization
				🟥		regulation of RNA splicing
				🟥		behavior
				🟥		macromolecule catabolic process
				🟦		regulation of postsynaptic membrane potential
				🟦		trans–synaptic signaling
				🟦		cognition
				🟦		regulation of biological quality
				🟦		neurogenesis
					🟦	monoatomic ion transmembrane transport
					🟦	intracellular calcium ion homeostasis
					🟦	glutamate receptor signaling pathway
					🟦	intracellular signaling cassette
					🟦	receptor localization to synapse
					🟦	rhythmic process
					🟦	protein kinase C–activating G protein–coupled receptor signaling pathway
					🟦	adenylate cyclase–inhibiting G protein–coupled glutamate receptor signaling pathway
					🟦	response to metal ion
					🟦	locomotion
					🟦	lipid phosphorylation
					🟦	startle response
					🟦	maintenance of postsynaptic density structure
					🟦	phospholipase C–activating G protein–coupled glutamate receptor signaling pathway
	🟥				🟦	calmodulin binding
	🟦					nitric–oxide synthase regulator activity
		🟥				brain–derived neurotrophic factor receptor activity
			🟥	🟥		protein binding
				🟦		transmitter–gated monoatomic ion channel activity
				🟦	🟦	PDZ domain binding
				🟦		axon guidance receptor activity
					🟦	monoatomic ion channel activity
					🟦	glutamate receptor activity
					🟦	neurotransmitter receptor activity involved in regulation of postsynaptic cytosolic calcium ion concentration
					🟦	structural constituent of synapse
					🟦	calcium ion binding
					🟦	structural constituent of postsynaptic density
					🟦	ATP–dependent diacylglycerol kinase activity
					🟦	calcium channel regulator activity
					🟦	ephrin receptor activity
					🟦	voltage–gated calcium channel activity involved in AV node cell action potential
					🟦	ryanodine–sensitive calcium–release channel activity

🟦 Enriched by downregulated DEGs
☐ Not significantly enriched
🟥 Enriched by upregulated DEGs

Figure 3.4.9 Excitatory neurons specific analysis.

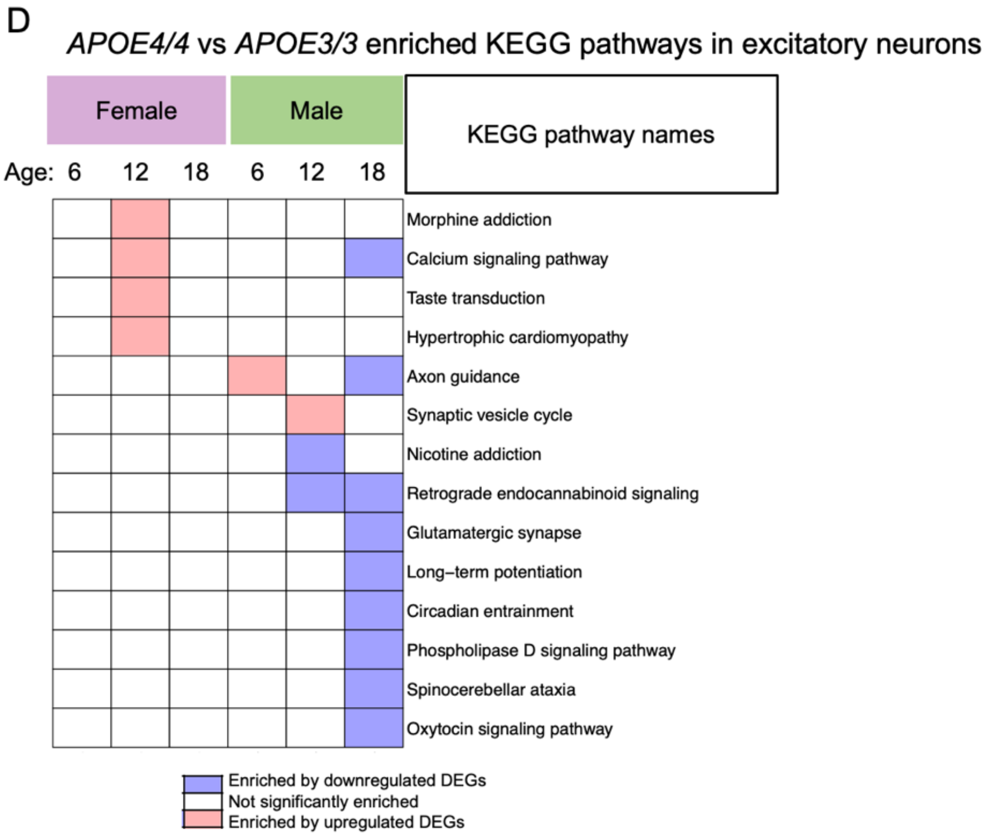


Figure 3.4.9 Excitatory neurons specific analysis. A. Upset plots showing *APOE4* vs *APOE3* DEG overlaps in excitatory neurons across sexes and ages, each condition further split by directions of gene regulations. B. Heatmap showing expression of a selected subset of DEGs in excitatory neurons that overlapped in males and females across age groups. C. Heatmap showing *APOE4* vs *APOE3* enriched GO terms in excitatory neurons stratified by age and sex. D. Heatmap showing *APOE4* vs *APOE3* enriched KEGG pathways in excitatory neurons stratified by age and sex. C,D. Colors indicate directions of regulation, with red labels for upregulated, blue labels for downregulated, and white labels non-significant in the associated groups.

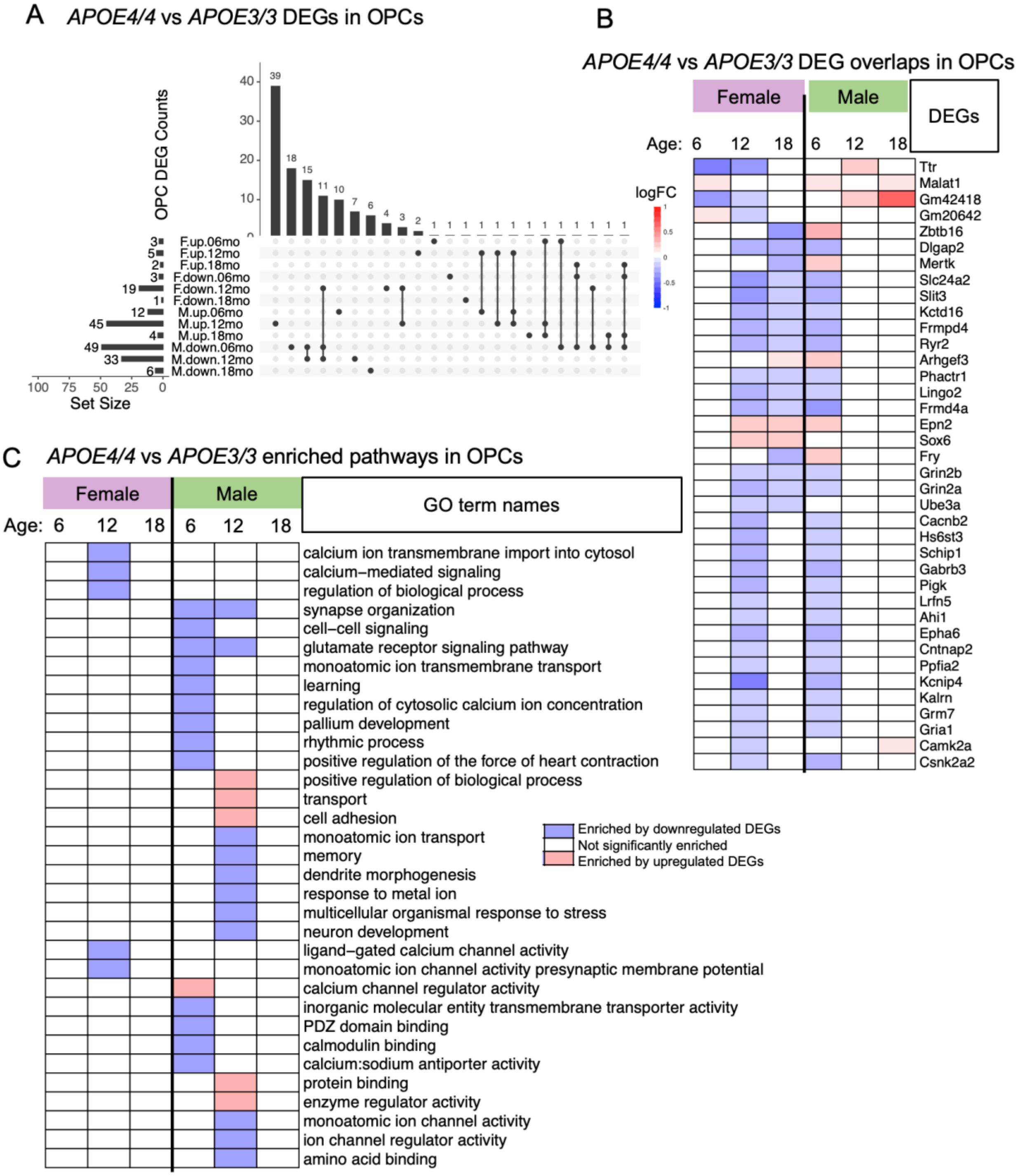


Figure 3.4.10 OPCs specific analysis.

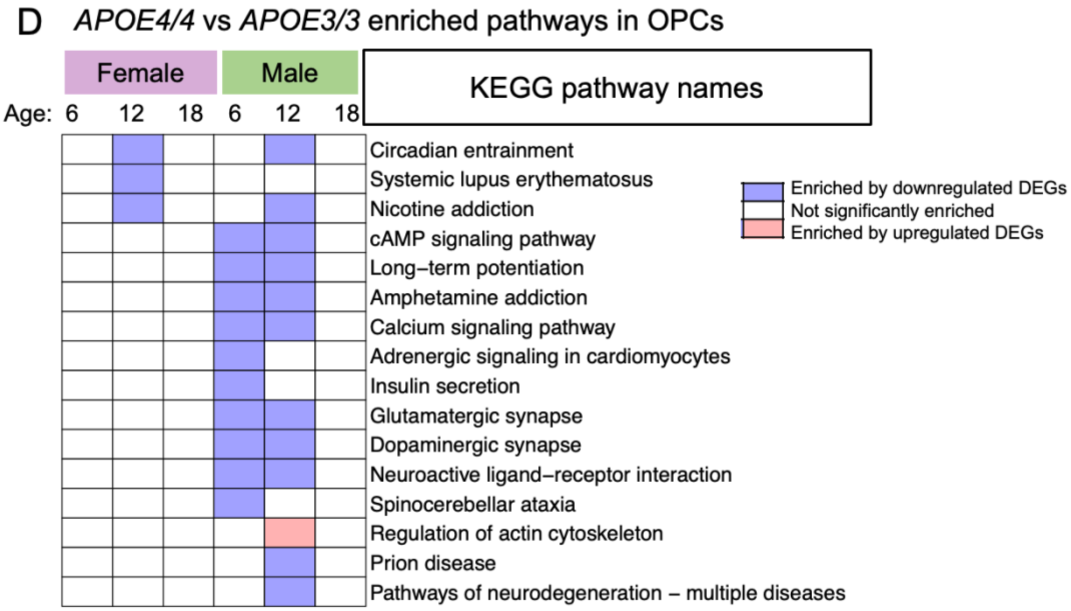
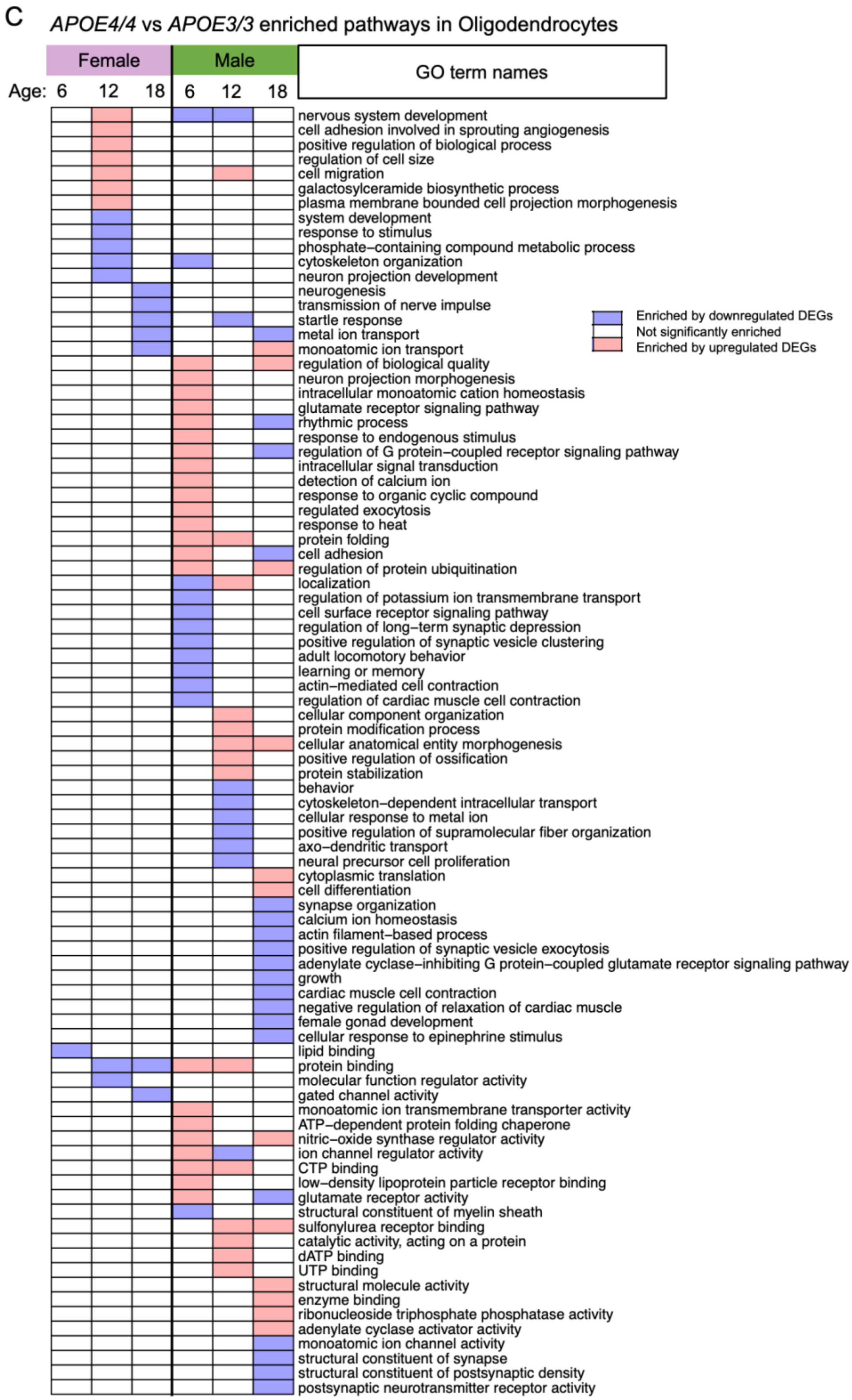


Figure 3.4.10 OPCs specific analysis. A. Upset plots showing *APOE4* vs *APOE3* DEG overlaps in OPCs across sexes and ages, each condition further split by directions of gene regulations. red line crosses gene intersects between females and males. B. Heatmap showing expression of DEGs in OPCs that overlapped in males and females across age groups. C. Heatmap showing *APOE4* vs *APOE3* enriched GO terms in OPCs stratified by age and sex. Colors indicate directions of regulation, with red labels for upregulated, blue labels for downregulated, and white labels non-significant in the associated groups. D. Heatmap showing *APOE4* vs *APOE3* enriched KEGG pathways in OPCs stratified by age and sex.



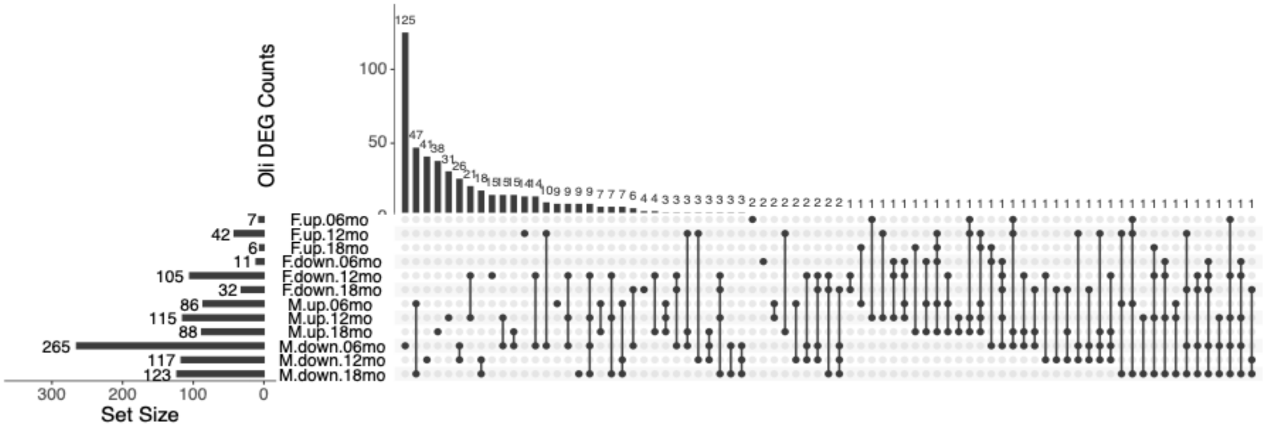
C *APOE4/4* vs *APOE3/3* enriched pathways in Oligodendrocytes

🟦 Enriched by downregulated DEGs
⬜ Not significantly enriched
🟥 Enriched by upregulated DEGs

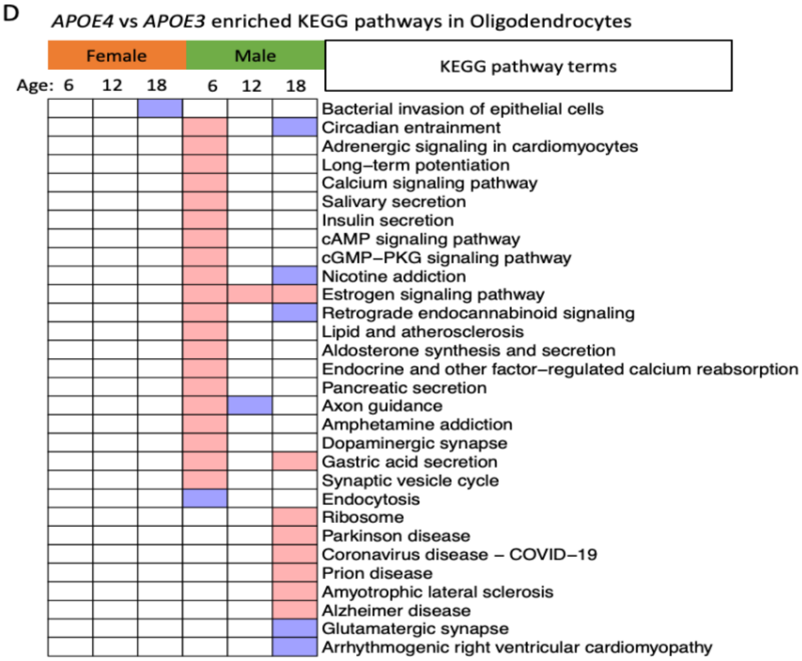
Female			Male			GO term names
Age: 6	12	18	6	12	18	
⬜	🟥	⬜	🟦	🟦	⬜	nervous system development
⬜	🟥	⬜	⬜	⬜	⬜	cell adhesion involved in sprouting angiogenesis
⬜	🟥	⬜	⬜	⬜	⬜	positive regulation of biological process
⬜	🟥	⬜	⬜	⬜	⬜	regulation of cell size
⬜	🟥	⬜	⬜	🟥	⬜	cell migration
⬜	🟥	⬜	⬜	⬜	⬜	galactosylceramide biosynthetic process
⬜	🟥	⬜	⬜	⬜	⬜	plasma membrane bounded cell projection morphogenesis
⬜	🟦	⬜	⬜	⬜	⬜	system development
⬜	🟦	⬜	⬜	⬜	⬜	response to stimulus
⬜	🟦	⬜	⬜	⬜	⬜	phosphate–containing compound metabolic process
⬜	🟦	⬜	🟦	⬜	⬜	cytoskeleton organization
⬜	🟦	⬜	⬜	⬜	⬜	neuron projection development
⬜	⬜	🟦	⬜	⬜	⬜	neurogenesis
⬜	⬜	🟦	⬜	⬜	⬜	transmission of nerve impulse
⬜	⬜	🟦	⬜	🟦	⬜	startle response
⬜	⬜	🟦	⬜	⬜	🟦	metal ion transport
⬜	⬜	🟦	⬜	⬜	🟥	monoatomic ion transport
⬜	⬜	⬜	🟥	⬜	🟥	regulation of biological quality
⬜	⬜	⬜	🟥	⬜	⬜	neuron projection morphogenesis
⬜	⬜	⬜	🟥	⬜	⬜	intracellular monoatomic cation homeostasis
⬜	⬜	⬜	🟥	⬜	⬜	glutamate receptor signaling pathway
⬜	⬜	⬜	🟥	⬜	🟦	rhythmic process
⬜	⬜	⬜	🟥	⬜	⬜	response to endogenous stimulus
⬜	⬜	⬜	🟥	⬜	🟦	regulation of G protein–coupled receptor signaling pathway
⬜	⬜	⬜	🟥	⬜	⬜	intracellular signal transduction
⬜	⬜	⬜	🟥	⬜	⬜	detection of calcium ion
⬜	⬜	⬜	🟥	⬜	⬜	response to organic cyclic compound
⬜	⬜	⬜	🟥	⬜	⬜	regulated exocytosis
⬜	⬜	⬜	🟥	⬜	⬜	response to heat
⬜	⬜	⬜	🟥	🟥	⬜	protein folding
⬜	⬜	⬜	🟥	⬜	🟦	cell adhesion
⬜	⬜	⬜	🟥	⬜	🟥	regulation of protein ubiquitination
⬜	⬜	⬜	🟦	🟥	⬜	localization
⬜	⬜	⬜	🟦	⬜	⬜	regulation of potassium ion transmembrane transport
⬜	⬜	⬜	🟦	⬜	⬜	cell surface receptor signaling pathway
⬜	⬜	⬜	🟦	⬜	⬜	regulation of long–term synaptic depression
⬜	⬜	⬜	🟦	⬜	⬜	positive regulation of synaptic vesicle clustering
⬜	⬜	⬜	🟦	⬜	⬜	adult locomotory behavior
⬜	⬜	⬜	🟦	⬜	⬜	learning or memory
⬜	⬜	⬜	🟦	⬜	⬜	actin–mediated cell contraction
⬜	⬜	⬜	🟦	⬜	⬜	regulation of cardiac muscle cell contraction
⬜	⬜	⬜	⬜	🟥	⬜	cellular component organization
⬜	⬜	⬜	⬜	🟥	⬜	protein modification process
⬜	⬜	⬜	⬜	🟥	🟥	cellular anatomical entity morphogenesis
⬜	⬜	⬜	⬜	🟥	⬜	positive regulation of ossification
⬜	⬜	⬜	⬜	🟥	⬜	protein stabilization
⬜	⬜	⬜	⬜	🟦	⬜	behavior
⬜	⬜	⬜	⬜	🟦	⬜	cytoskeleton–dependent intracellular transport
⬜	⬜	⬜	⬜	🟦	⬜	cellular response to metal ion
⬜	⬜	⬜	⬜	🟦	⬜	positive regulation of supramolecular fiber organization
⬜	⬜	⬜	⬜	🟦	⬜	axo–dendritic transport
⬜	⬜	⬜	⬜	🟦	⬜	neural precursor cell proliferation
⬜	⬜	⬜	⬜	⬜	🟥	cytoplasmic translation
⬜	⬜	⬜	⬜	⬜	🟥	cell differentiation
⬜	⬜	⬜	⬜	⬜	🟦	synapse organization
⬜	⬜	⬜	⬜	⬜	🟦	calcium ion homeostasis
⬜	⬜	⬜	⬜	⬜	🟦	actin filament–based process
⬜	⬜	⬜	⬜	⬜	🟦	positive regulation of synaptic vesicle exocytosis
⬜	⬜	⬜	⬜	⬜	🟦	adenylate cyclase–inhibiting G protein–coupled glutamate receptor signaling pathway
⬜	⬜	⬜	⬜	⬜	🟦	growth
⬜	⬜	⬜	⬜	⬜	🟦	cardiac muscle cell contraction
⬜	⬜	⬜	⬜	⬜	🟦	negative regulation of relaxation of cardiac muscle
⬜	⬜	⬜	⬜	⬜	🟦	female gonad development
⬜	⬜	⬜	⬜	⬜	🟦	cellular response to epinephrine stimulus
🟦	⬜	⬜	⬜	⬜	⬜	lipid binding
⬜	🟦	🟦	🟥	🟥	⬜	protein binding
⬜	🟦	⬜	⬜	⬜	⬜	molecular function regulator activity
⬜	⬜	🟦	⬜	⬜	⬜	gated channel activity
⬜	⬜	⬜	🟥	⬜	⬜	monoatomic ion transmembrane transporter activity
⬜	⬜	⬜	🟥	⬜	⬜	ATP–dependent protein folding chaperone
⬜	⬜	⬜	🟥	⬜	🟥	nitric–oxide synthase regulator activity
⬜	⬜	⬜	🟥	🟦	⬜	ion channel regulator activity
⬜	⬜	⬜	🟥	🟥	⬜	CTP binding
⬜	⬜	⬜	🟥	⬜	⬜	low–density lipoprotein particle receptor binding
⬜	⬜	⬜	🟥	⬜	🟦	glutamate receptor activity
⬜	⬜	⬜	🟦	⬜	⬜	structural constituent of myelin sheath
⬜	⬜	⬜	⬜	🟥	🟥	sulfonylurea receptor binding
⬜	⬜	⬜	⬜	🟥	⬜	catalytic activity, acting on a protein
⬜	⬜	⬜	⬜	🟥	⬜	dATP binding
⬜	⬜	⬜	⬜	🟥	⬜	UTP binding
⬜	⬜	⬜	⬜	⬜	🟥	structural molecule activity
⬜	⬜	⬜	⬜	⬜	🟥	enzyme binding
⬜	⬜	⬜	⬜	⬜	🟥	ribonucleoside triphosphate phosphatase activity
⬜	⬜	⬜	⬜	⬜	🟥	adenylate cyclase activator activity
⬜	⬜	⬜	⬜	⬜	🟦	monoatomic ion channel activity
⬜	⬜	⬜	⬜	⬜	🟦	structural constituent of synapse
⬜	⬜	⬜	⬜	⬜	🟦	structural constituent of postsynaptic density
⬜	⬜	⬜	⬜	⬜	🟦	postsynaptic neurotransmitter receptor activity

Figure 3.4.11 Oligodendrocyte specific analysis

A



D



Female 6	Female 12	Female 18	Male 6	Male 12	Male 18	KEGG pathway terms
		🟦				Bacterial invasion of epithelial cells
			🟥		🟦	Circadian entrainment
			🟥			Adrenergic signaling in cardiomyocytes
			🟥			Long−term potentiation
			🟥			Calcium signaling pathway
			🟥			Salivary secretion
			🟥			Insulin secretion
			🟥			cAMP signaling pathway
			🟥			cGMP−PKG signaling pathway
			🟥		🟦	Nicotine addiction
			🟥	🟥	🟥	Estrogen signaling pathway
			🟥		🟦	Retrograde endocannabinoid signaling
			🟥			Lipid and atherosclerosis
			🟥			Aldosterone synthesis and secretion
			🟥			Endocrine and other factor−regulated calcium reabsorption
			🟥			Pancreatic secretion
			🟥	🟦		Axon guidance
			🟥			Amphetamine addiction
			🟥			Dopaminergic synapse
			🟥		🟥	Gastric acid secretion
			🟥			Synaptic vesicle cycle
			🟦			Endocytosis
					🟥	Ribosome
					🟥	Parkinson disease
					🟥	Coronavirus disease − COVID−19
					🟥	Prion disease
					🟥	Amyotrophic lateral sclerosis
					🟥	Alzheimer disease
					🟦	Glutamatergic synapse
					🟦	Arrhythmogenic right ventricular cardiomyopathy

Figure 3.4.11 Oligodendrocyte specific analysis. A. Upset plots showing *APOE4* vs *APOE3* DEG overlaps in oligodendrocytes across sexes and ages, each condition further split by directions of gene regulations. red line crosses gene intersects between females and males. C, D Heatmap showing *APOE4* vs *APOE3* enriched GO terms (C) KEGG pathways (D) in oligodendrocytes stratified by age and sex. Colors indicate directions of regulation, with red labels for upregulated, blue labels for downregulated, and white labels non-significant in the associated groups.

CHAPTER 4: CONCLUSION

Alzheimer's disease (AD), the most common form of dementia, is an age-related neurodegenerative disorder, and no preventative or curative treatments currently exist[8]. In aging populations, AD and related dementia affects over 50 million elderly people worldwide[86]. By 2050, it is anticipated that this number will triple due to the high aging population that will create a significant load on healthcare systems[87]. The costs associated with managing AD, including medical care, long-term care, and informal caregiving, are enormous[88]. In 2020, the global cost of dementia, including AD, was estimated to be over $1 trillion, a figure that is expected to double by 2030[4]. AD leads to progressive cognitive decline, memory loss, and functional impairment, ultimately resulting in complete dependency on caregivers[52]. This decline severely affects the quality of life for both patients and their families. AD is a global crisis due to its increasing prevalence, economic burden, profound impact on quality of life, and the current lack of effective treatments.

AD is effectively incurable because current treatment options available to patients have a negligible impact on the severity and progression of the disease. Despite significant research efforts, there are only a handful of treatment options[6]. Majority of existing therapies primarily focus on managing symptoms rather than slowing or halting disease progression. The failure of numerous clinical trials highlights the complexity of AD and the urgent need for innovative therapeutic approaches. Additionally, AD is often diagnosed in its later stages when irreversible brain damage has already occurred[19]. This delay in diagnosis limits the effectiveness of potential interventions and underscores the need for better diagnostic tools and early detection methods.

Although our understanding of the molecular mechanisms underpinning the progression of AD has increased steadily over the past decades, the precise etiology of the disease remains elusive. As our understanding of disease mechanisms and biological networks grows, the design of drugs that strategically exploit multi-target effects is likely to become a key strategy in developing more effective and safer therapies for complex diseases, such as AD. A multi-target combination therapy might be more effective at treating AD by rectifying the heterogeneous disease mechanisms across various disease-implicated cell types. Besides the multifactorial pathogenesis aspect of the disease, the enormous heterogeneity in AD might require precision medicine. By focusing on key risk factors such as age, APOE genotype, sex, we can identify individuals at higher risk of developing AD before the onset of clinical symptoms, allowing for earlier intervention. Moreover, a deep understanding of the molecular pathology associated with these risk factors—such as the alterations in gene expression and protein function—can illuminate the underlying biological networks that drive disease progression. By mapping these networks, we can identify potential therapeutic targets for gene expression correction specific to the risk-associated patient populations, offering a more precise and personalized approach to treatment. This approach not only aids in the early diagnosis of AD but also paves the way for developing network-based therapies that can modulate the expression of key genes, thereby addressing the multifactorial nature of AD and potentially altering its trajectory. Integrating risk factor analysis with molecular insights thus holds significant potential for improving both the early diagnosis and treatment of AD.

With these objectives in mind, in chapter 2, we embarked on a pioneering exploration of the therapeutic potential of combination therapy, specifically targeting cell-type-specific disease signatures at the transcriptomic level across multiple cell types implicated in AD. Despite over 2,700 clinical trials primarily targeting single core pathologies were initiated in the past two decades, only two received FDA approval[9]. The limited success of single-target approaches underscores the need for strategies addressing AD's heterogeneous pathologies across various cell types, such as synaptic loss, neuronal death, neuroinflammation, cholesterol homeostasis alteration, energy metabolism deficit, dysfunctional phagocytosis, and demyelination[10]. To address this complexity, we designed a cell-type-directed combination therapy to correct cell-type-specific dysregulated gene networks in both neuronal and glial cells. This network-correction approach aims to restore normal cellular function by targeting multiple AD-related pathways simultaneously. By reversing the expression of hundreds of disease signature genes within distinct disease-relevant cell types, this strategy offers a more comprehensive and effective therapeutic solution.

While systems biology approaches and single-cell data mining have been employed for computational drug discovery in AD[89,90], they often produce extensive lists of candidates without adequate validation or prioritization, limiting clinical translation. We addressed this issue leveraging real-world evidence from the UC-wide EMR to investigate AD diagnosis in patients with drug exposure. Among the predicted drug candidates with sufficient patient representation, most were associated with decreased relative risk, suggesting potential beneficial effects in humans. Based on their cell-type-specific targets, we prioritized letrozole (neurons) and irinotecan (glial cells) as a combination therapy among all EMR-validated drugs. Additionally, both drugs effectively cross the blood-brain barrier[91,92], ensuring they reach the brain and perform their intended therapeutic functions. Furthermore, previous epidemiological research indicated a reduced dementia risk among breast cancer patients treated with letrozole[93] and decreased AD risk among colorectal cancer survivors treated with irinotecan[94]. These findings support the potential efficacy of letrozole and irinotecan as components of a combination therapy for AD.

A thorough investigation of the network-correcting effects of the combination therapy was conducted with snRNA-seq analysis of hippocampal samples from combination-treated and vehicle-treated AD mice carrying Aβ and tau pathologies. In addition to rescuing memory deficits, hippocampal atrophy, neuronal loss, neuroinflammation, and tau and Aβ pathologies, we found many AD-related pathways enriched by the combination treatment-reversed genes. These include axonogenesis, neuron projections, synaptic activities, and regulation of protein phosphorylation in neurons; cholesterol metabolism and regulation of long-term synaptic potentiation in astrocytes; synaptic organization and cell growth regulation in microglia; and regulation of neuronal projections and synaptic signaling in oligodendrocytes and OPCs. These observations of treatment-reversed AD-associated biological processes supports our hypothesis that a targeted, multifaceted therapeutic approach provides AD-specific beneficial effects possibly through rectifying multiple AD pathological processes in a cell-type-specific manner. Further investigation is warranted to elucidate the precise mechanisms and key pathways targeted by these drugs.

Despite these promising findings, several limitations of this study should be noted. First, since CMAP drug profiles were measured in cancer cells and may not accurately reflect brain tissue profiles; they were used to screen drug candidates and require validations in neurological models. Relevant drug perturbation databases on different brain cell types are currently unavailable. Second, reliance on animal models, though necessary, may not fully represent human biology. Significant sex differences to treatment were observed in behavior tests and pathology analyses, with male mice generally responding better behaviorally to all treatments, though the pattern varied across pathology markers. Although sex differences in AD are well-documented and the impact of sex hormones is inconclusive[95], no significant difference in AD prevalence after drug exposures was observed between sexes in the human data from the UC-wide EMR sex-stratified analysis. This discrepancy suggests the observed sex differences may be specific to the mouse models, as both 5xFAD and PS19 have shown strong sex difference in prior studies[96,97]. Additionally, EMR analysis presents potential caveats, as data tend to be sparse and are not collected with specific research in mind. Two of the ten drug candidates screened using EMR data, both used in neurological disorders, showed significantly higher relative risk scores for AD with exposure, suggesting detrimental effects. However, these neurological conditions are AD comorbidities, increasing the likelihood that AD patients were receiving these drugs. Despite efforts to propensity-match, the control populations may accurately represent the drug-exposed populations if relevant diagnoses are missing. Therefore, rather than relying solely on relative risk scores to interpret drug effects, we used EMR results to prioritize candidates with the best scores for further experimental validation.

By validating the therapeutic benefits of such a combination, this research serves as a proof-of-concept, demonstrating the feasibility of a multi-cell-type gene network correction therapy that may more effectively ameliorate AD-related pathologies. While this treatment method accounts for the multifactorial nature of AD, recognizing the vast heterogeneity among AD patients suggests that effective treatment strategies will likely involve artificial intelligence (AI)-driven precision medicine. Therefore, in Chapter 3, we systematically characterized the cellular and molecular alterations induced by major AD risk factors, including age, sex, and *APOE* genotypes, to gain a comprehensive understanding of the functional changes associated with each individual factor, as well as the convergent effects of multiple risk factors in a cell-type-specific context.

With this comprehensive survey study, we found sex, out of all three risk factors, is consistently the strongest variance contributing factor in cell type abundance, specifically for inhibitory neurons, astrocytes, and OPCs. Additionally, sex significantly contributes to variance in abundance of neuronal subtypes in the hippocampus. These data suggest the importance of taking the female sex in consideration when studying disease mechanisms. Furthermore, we observed age-dependent sex difference in cell-cell communications comparing *APOE4* vs *APOE3* samples across cell types. Increased communication comparing *APOE4* vs *APOE3* samples was observed in female with aging, while decreased communication by *APOE4* allele was observed in males with aging. Lastly, we performed differential gene expression and enriched pathways analysis of *APOE4* vs *APOE3* samples and observed age-dependent sex-bias in molecular alterations by *APOE* genotypes across all cell types. This rigorous profiling not only facilitates a better understanding of the intricate heterogeneity observed in AD patient populations but also establishes a robust foundation for the future development of targeted network correction therapies within a precision medicine framework. This large-scale transcriptomic dataset enables the identification of nuanced patterns that reveal molecular and functional deficits in specific subgroups within the AD population, allowing for the design of tailored therapies targeting these specific patterns of alterations. By integrating insights from diverse cellular and molecular landscapes, this study advances the prospect of designing more precise and effective therapeutic interventions tailored to the complex and multifactorial nature of AD.

In summary, the first study introduces a transformative approach to AD drug discovery by focusing on cell-type-specific interventions and predicting multi-target drug combinations based on human-derived data. Utilizing extensive omic data from post-mortem human brains and clinical records from millions of diverse patients, our methodology provides a robust framework for identifying effective therapeutic agents through real-world evidence. Preclinical validation in mouse models showed that the proposed drug combination significantly ameliorated memory deficits and AD-related pathologies. Although treatment reversed AD-specific gene network signatures, further investigation is needed to elucidate the precise mechanistic targets to advance it as a novel AD therapeutic.

The results of the second project have significant implications for the advancement of precision medicine in AD. By comprehensively characterizing the molecular influences of key AD risk factors—such as age, *APOE4* genotype, and sex—using AD mouse models, the study has identified substantial variations in cell type abundance and gene expression patterns, particularly driven by sex differences and the interplay between age and APOE genotype. These findings not only deepen our understanding of the molecular underpinnings of AD but also create a rich dataset that can be leveraged for future applications of the network correction method. Importantly, this dataset allows for the mapping of distinct risk profiles, facilitating the identification of tailored therapeutic interventions that align with the unique transcriptomic signatures of individuals. Ultimately, this work enhances the potential for developing more personalized and effective treatment strategies for AD, addressing the disease's complex and multifactorial nature through precision medicine.

Collectively, these two complementary projects underscore the potential of a cell-type-directed, network-correcting treatment strategy within a precision medicine framework. This work lays the groundwork for advancing precision medicine by leveraging AI and large-scale omics datasets. By integrating advanced computational techniques with multi-omic, clinical, and epidemiological databases, future initiatives can develop AI-driven systems capable of designing tailored treatments based on the molecular signatures and clinical profiles of specific subpopulations or individuals. This approach directly addresses the multifactorial nature of AD pathology and heterogeneity observed among AD patients, thereby facilitating the development of more precise, predictive, and effective therapeutic interventions.

REFERENCES

1. Masters, C. L. *et al.* Alzheimer’s disease. *Nature Reviews Disease Primers* **1**, 1–18 (2015).
2. Prince, M. J., Comas-Herrera, A., Knapp, M., Guerchet, M. M. & Karagiannidou, M. World Alzheimer Report 2016 - Improving healthcare for people living with dementia: Coverage, quality and costs now and in the future. (2016).
3. Hebert, L. E., Weuve, J., Scherr, P. A. & Evans, D. A. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* **80**, 1778–1783 (2013).
4. World Health Organization. Global status report on the public health response to dementia. https://www.who.int/publications-detail-redirect/9789240033245.
5. Yiannopoulou, K. G. & Papageorgiou, S. G. Current and future treatments for Alzheimer’s disease. *Ther Adv Neurol Disord* **6**, 19–33 (2013).
6. How Is Alzheimer’s Disease Treated? *National Institute on Aging* http://www.nia.nih.gov/health/how-alzheimers-disease-treated.
7. Valiukas, Z. *et al.* Immunotherapies for Alzheimer’s Disease—A Review. *Vaccines (Basel)* **10**, 1527 (2022).
8. Frozza, R. L., Lourenco, M. V. & De Felice, F. G. Challenges for Alzheimer’s Disease Therapy: Insights from Novel Mechanisms Beyond Memory Defects. *Front Neurosci* **12**, (2018).
9. Kim, C. K. *et al.* Alzheimer’s Disease: Key Insights from Two Decades of Clinical Trial Failures. *J Alzheimers Dis* **87**, 83–100 (2022).
10. Toledano-Díaz, A., Álvarez, M. I. & Toledano, A. The relationships between neuroglial alterations and neuronal changes in Alzheimer’s disease, and the related controversies I: Gliopathogenesis and glioprotection. *J Cent Nerv Syst Dis* **14**, 11795735221128703 (2022).
11. Blumenfeld, J., Yip, O., Kim, M. J. & Huang, Y. Cell type-specific roles of APOE4 in Alzheimer disease. *Nat Rev Neurosci* **25**, 91–110 (2024).

12. Bloom, G. S. Amyloid-β and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol* **71**, 505–508 (2014).

13. Hansen, D. V., Hanson, J. E. & Sheng, M. Microglia in Alzheimer's disease. *J Cell Biol* **217**, 459–472 (2018).

14. Cai, Z. & Xiao, M. Oligodendrocytes and Alzheimer's disease. *Int J Neurosci* **126**, 97–104 (2016).

15. González-Reyes, R. E., Nava-Mesa, M. O., Vargas-Sánchez, K., Ariza-Salamanca, D. & Mora-Muñoz, L. Involvement of Astrocytes in Alzheimer's Disease from a Neuroinflammatory and Oxidative Stress Perspective. *Front Mol Neurosci* **10**, (2017).

16. Barnes, L. L. *et al.* Sex Differences in the Clinical Manifestations of Alzheimer Disease Pathology. *Archives of General Psychiatry* **62**, 685–691 (2005).

17. Guerreiro, R. J., Gustafson, D. R. & Hardy, J. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. *Neurobiol Aging* **33**, 437–456 (2012).

18. Liyanage, S. I., Santos, C. & Weaver, D. F. The hidden variables problem in Alzheimer's disease clinical trial design. *Alzheimers Dement (N Y)* **4**, 628–635 (2018).

19. Reitz, C., Rogaeva, E. & Beecham, G. W. Late-onset vs nonmendelian early-onset Alzheimer disease. *Neurol Genet* **6**, e512 (2020).

20. Dubal, D. B. Sex difference in Alzheimer's disease: An updated, balanced and emerging perspective on differing vulnerabilities. *Handb Clin Neurol* **175**, 261–273 (2020).

21. Buckley, R. F. *et al.* Sex Differences in the Association of Global Amyloid and Regional Tau Deposition Measured by Positron Emission Tomography in Clinically Normal Older Adults. *JAMA Neurol* **76**, 542–551 (2019).

22. Zhao, N. *et al.* Alzheimer's Risk Factors Age, APOE Genotype, and Sex Drive Distinct Molecular Pathways. *Neuron* **106**, 727-742.e6 (2020).

23. Barron, A. M. & Pike, C. J. Sex hormones, aging, and Alzheimer's disease. *Front Biosci (Elite Ed)* **4**, 976–997 (2012).

24. Rabinovici, G. D. Late-onset Alzheimer Disease. *Continuum (Minneap Minn)* **25**, 14–33 (2019).
25. Medicare, I. of M. (US) C. to D. a S. for Q. R. and A. in & Lohr, K. N. The Elderly Population. in *Medicare: A Strategy for Quality Assurance: Volume 1* (National Academies Press (US), 1990).
26. Tanzi, R. E. The Genetics of Alzheimer Disease. *Cold Spring Harb Perspect Med* **2**, a006296 (2012).
27. Najm, R., Jones, E. A. & Huang, Y. Apolipoprotein E4, inhibitory network dysfunction, and Alzheimer's disease. *Mol Neurodegeneration* **14**, 1–13 (2019).
28. Belloy, M. E. *et al.* APOE Genotype and Alzheimer Disease Risk Across Age, Sex, and Population Ancestry. *JAMA Neurology* **80**, 1284–1294 (2023).
29. Shi, Y. *et al.* ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature* **549**, 523–527 (2017).
30. Benson, G. S. *et al.* Don't forget about tau: the effects of ApoE4 genotype on Alzheimer's disease cerebrospinal fluid biomarkers in subjects with mild cognitive impairment—data from the Dementia Competence Network. *J Neural Transm (Vienna)* **129**, 477–486 (2022).
31. H, M. *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, (2023).
32. Davis, E. J. *et al.* A second X chromosome contributes to resilience in a mouse model of Alzheimer's disease. *Sci Transl Med* **12**, (2020).
33. Mielke, M. M., Vemuri, P. & Rocca, W. A. Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences. *Clin Epidemiol* **6**, 37–48 (2014).
34. Oveisgharan, S. *et al.* Sex differences in Alzheimer's disease and common neuropathologies of aging. *Acta Neuropathol* **136**, 887–900 (2018).
35. Hohman, T. J. *et al.* Sex-Specific Association of Apolipoprotein E With Cerebrospinal Fluid Levels of Tau. *JAMA Neurol* **75**, 989–998 (2018).
36. Riedel, B. C., Thompson, P. M. & Brinton, R. D. Age, APOE and Sex: Triad of Risk of Alzheimer's Disease. *J Steroid Biochem Mol Biol* **160**, 134–147 (2016).

37. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).

38. Zhou, Y. *et al.* Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med* **26**, 131–142 (2020).

39. Lau, S.-F., Cao, H., Fu, A. K. Y. & Ip, N. Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *PNAS* **117**, 25800–25809 (2020).

40. Taubes, A. *et al.* Experimental and real-world evidence supporting the computational repurposing of bumetanide for APOE4-related Alzheimer's disease. *Nat Aging* **1**, 932–947 (2021).

41. Sirota, M. *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* **3**, 96ra77 (2011).

42. Lamb, J. *et al.* The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* **313**, 1929–1935 (2006).

43. Grabowska, M. E., Huang, A., Wen, Z., Li, B. & Wei, W.-Q. Drug repurposing for Alzheimer's disease from 2012–2022—a 10-year literature review. *Front Pharmacol* **14**, 1257700 (2023).

44. Stewart, A. K. Medicine. How thalidomide works against cancer. *Science* **343**, 256–257 (2014).

45. Hubsher, G., Haider, M. & Okun, M. S. Amantadine: the journey from fighting flu to treating Parkinson disease. *Neurology* **78**, 1096–1099 (2012).

46. Appleby, B. S., Nacopoulos, D., Milano, N., Zhong, K. & Cummings, J. L. A Review: Treatment of Alzheimer's Disease Discovered in Repurposed Agents. *DEM* **35**, 1–22 (2013).

47. Tang, A. S. *et al.* Deep phenotyping of Alzheimer's disease leveraging electronic medical records identifies sex-specific clinical associations. *Nat Commun* **13**, 675 (2022).

48. Xu, H., Li, J., Jiang, X. & Chen, Q. Electronic Health Records for Drug Repurposing: Current Status, Challenges, and Future Directions. *Clin Pharmacol Ther* **107**, 712–714 (2020).

49. Alzheimer's Association 2024 Alzheimer's Disease Facts and Figures.

50. What Is Alzheimer's Disease? *National Institute on Aging* https://www.nia.nih.gov/health/alzheimers-and-dementia/what-alzheimers-disease (2021).
51. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **7**, e105–e125 (2022).
52. Alzheimer's Disease Fact Sheet. *National Institute on Aging* http://www.nia.nih.gov/health/alzheimers-disease-fact-sheet.
53. Dementia. https://www.who.int/news-room/fact-sheets/detail/dementia.
54. Cheung, C., Goh, Y. T., Zhang, J., Wu, C. & Guccione, E. Modeling cerebrovascular pathophysiology in amyloid-β metabolism using neural-crest-derived smooth muscle cells. *Cell Rep* **9**, 391–401 (2014).
55. Choi, Y. *et al.* Clozapine Improves Memory Impairment and Reduces Aβ Level in the Tg-APPswe/PS1dE9 Mouse Model of Alzheimer's Disease. *Mol Neurobiol* **54**, 450–460 (2017).
56. Gerring, Z. F., Gamazon, E. R., White, A. & Derks, E. M. Integrative Network-Based Analysis Reveals Gene Networks and Novel Drug Repositioning Candidates for Alzheimer Disease. *Neurology Genetics* **7**, e622 (2021).
57. Rodriguez, S. *et al.* Machine learning identifies candidates for drug repurposing in Alzheimer's disease. *Nat Commun* **12**, 1033 (2021).
58. 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. https://www.nature.com/articles/s43587-021-00138-z.
59. Pajai, S., Potdar, J., Gopal, U. & Banait, T. A Review on the Use of Letrozole in Female and Male Infertility. *Cureus* **14**, e31291.
60. Fujita, K., Kubota, Y., Ishida, H. & Sasaki, Y. Irinotecan, a key chemotherapeutic drug for metastatic colorectal cancer. *World J Gastroenterol* **21**, 12234–12248 (2015).

61. Tau P301S (Line PS19) | ALZFORUM. https://www.alzforum.org/research-models/tau-p301s-line-ps19.
62. 5xFAD (B6SJL) | ALZFORUM. https://www.alzforum.org/research-models/5xfad-b6sjl.
63. Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479–486 (1991).
64. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol* **112**, 389–404 (2006).
65. Leung, L. *et al.* Apolipoprotein E4 Causes Age- and Sex-Dependent Impairments of Hilar GABAergic Interneurons and Learning and Memory Deficits in Mice. *PLoS ONE* **7**, (2012).
66. Knox, C. *et al.* DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res* **52**, D1265–D1275 (2024).
67. Lamb, H. M. & Adkins, J. C. Letrozole. A review of its use in postmenopausal women with advanced breast cancer. *Drugs* **56**, 1125–1140 (1998).
68. Saul, A., Sprenger, F., Bayer, T. A. & Wirths, O. Accelerated tau pathology with synaptic and neuronal loss in a novel triple transgenic mouse model of Alzheimer's disease. *Neurobiology of Aging* **34**, 2564–2573 (2013).
69. Stancu, I.-C. *et al.* Tauopathy contributes to synaptic and cognitive deficits in a murine model for Alzheimer's disease. *The FASEB Journal* **28**, 2620–2631 (2014).
70. Morris water maze: procedures for assessing spatial and related forms of learning and memory | Nature Protocols. https://www.nature.com/articles/nprot.2006.116.
71. Integrative analysis in Seurat v5 • Seurat. https://satijalab.org/seurat/articles/seurat5_integration.
72. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).

73. Mou, T., Deng, W., Gu, F., Pawitan, Y. & Vu, T. N. Reproducibility of Methods to Detect Differentially Expressed Genes from Single-Cell RNA Sequencing. *Front. Genet.* **10**, (2020).

74. Alzheimer's Association. 2023 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **19**, 1598–1695 (2023).

75. Miramontes, S. *et al.* Alzheimer's disease as a women's health challenge: a call for action on integrative precision medicine approaches. *NPJ Womens Health* **2**, 17 (2024).

76. Belonwu, S. A. *et al.* Bioinformatics Analysis of Publicly Available Single-Nuclei Transcriptomics Alzheimer's Disease Datasets Reveals APOE Genotype-Specific Changes Across Cell Types in Two Brain Regions. *Front Aging Neurosci* **14**, 749991 (2022).

77. Neu, S. C. *et al.* Apolipoprotein E Genotype and Sex Risk Factors for Alzheimer Disease: A Meta-analysis. *JAMA Neurol* **74**, 1178–1189 (2017).

78. Deming, Y. *et al.* Sex-specific genetic predictors of Alzheimer's disease biomarkers. *Acta Neuropathol* **136**, 857–872 (2018).

79. Koran, M. E. I., Wagener, M. & Hohman, T. J. Sex Differences in the Association between AD Biomarkers and Cognitive Decline. *Brain Imaging Behav* **11**, 205–213 (2017).

80. Belonwu, S. A. *et al.* Sex-Stratified Single-Cell RNA-Seq Analysis Identifies Sex-Specific and Cell Type-Specific Transcriptional Responses in Alzheimer's Disease Across Two Brain Regions. *Mol Neurobiol* **59**, 276–293 (2022).

81. Sala Frigerio, C. *et al.* The Major Risk Factors for Alzheimer's Disease: Age, Sex, and Genes Modulate the Microglia Response to Aβ Plaques. *Cell Rep* **27**, 1293-1306.e6 (2019).

82. Stephen, T. L. *et al.* APOE genotype and sex affect microglial interactions with plaques in Alzheimer's disease mice. *Acta Neuropathol Commun* **7**, 82 (2019).

83. Neuroligin-1 in brain and CSF of neurodegenerative disorders: investigation for synaptic biomarkers - PubMed. https://pubmed.ncbi.nlm.nih.gov/33522967/.

84. Woo, R.-S., Lee, J.-H., Yu, H.-N., Song, D.-Y. & Baik, T.-K. Expression of ErbB4 in the neurons of

Alzheimer’s disease brain and APP/PS1 mice, a model of Alzheimer’s disease. *Anat Cell Biol* **44**, 116–127 (2011).

85. Inference and analysis of cell-cell communication using CellChat | Nature Communications. https://www.nature.com/articles/s41467-021-21246-9.
86. Plassman, B. L. *et al.* Prevalence of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology* **29**, 125–132 (2007).
87. Akushevich, I., Yashkin, A. P., Kravchenko, J. & Kertai, M. D. Chemotherapy and the Risk of Alzheimer’s Disease in Colorectal Cancer Survivors: Evidence From the Medicare System. *JCO Oncol Pract* **17**, e1649–e1659 (2021).
88. Economic Burden of Alzheimer Disease and Managed Care Considerations. *AJMC* https://www.ajmc.com/view/economic-burden-of-alzheimer-disease-and-managed-care-considerations.
89. Peyton, M. E. Single-cell Approach to Repurposing of Drugs for Alzheimer’s Disease. (2023).
90. Parolo, S., Mariotti, F., Bora, P., Carboni, L. & Domenici, E. Single-cell-led drug repurposing for Alzheimer’s disease. *Sci Rep* **13**, 222 (2023).
91. Dave, N., Gudelsky, G. A. & Desai, P. B. The pharmacokinetics of letrozole in brain and brain tumor in rats with orthotopically implanted C6 glioma, assessed using intracerebral microdialysis. *Cancer Chemother Pharmacol* **72**, 349–357 (2013).
92. Goldwirt, L., Beccaria, K., Carpentier, A., Farinotti, R. & Fernandez, C. Irinotecan and temozolomide brain distribution: a focus on ABCB1. *Cancer Chemother Pharmacol* **74**, 185–193 (2014).
93. Branigan, G. L., Soto, M., Neumayer, L., Rodgers, K. & Brinton, R. D. Association Between Hormone-Modulating Breast Cancer Therapies and Incidence of Neurodegenerative Outcomes for Women With Breast Cancer. *JAMA Network Open* **3**, e201541 (2020).
94. Stenger, M. Use of Chemotherapy and Risk of Alzheimer’s Disease in Colorectal Cancer Survivors.

https://ascopost.com/news/march-2021/use-of-chemotherapy-and-risk-of-alzheimer-s-disease-in-colorectal-cancer-survivors/.

95. Guo, L., Zhong, M. B., Zhang, L., Zhang, B. & Cai, D. Sex Differences in Alzheimer’s Disease: Insights From the Multiomics Landscape. *Biol Psychiatry* **91**, 61–71 (2022).
96. Sun, Y. *et al.* The behavioural and neuropathologic sexual dimorphism and absence of MIP-3α in tau P301S mouse model of Alzheimer’s disease. *Journal of Neuroinflammation* **17**, 72 (2020).
97. Sil, A. *et al.* Sex Differences in Behavior and Molecular Pathology in the 5XFAD Model. *J Alzheimers Dis* **85**, 755–778 (2022).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Yaqiao Li
4946C50B5C0F428...
Author Signature

8/31/2024
Date