



Systems biology approaches to unravel the molecular and genetic architecture of Alzheimer's disease and related tauopathies

Emily Miyoshi^{a,b,1}, Samuel Morabito^{b,c,1}, Vivek Swarup^{a,b,*}

^a Department of Neurobiology and Behavior, University of California, Irvine, CA 92697, USA

^b Institute for Memory Impairments and Neurological Disorders (MIND), University of California, Irvine, CA 92697, USA

^c Mathematical, Computational and Systems Biology (MCSB) Program, University of California, Irvine, CA 92697, USA

ARTICLE INFO

Keywords:

Tauopathy
Alzheimer's disease
Genetics
Transcriptomics
Epigenetics
Network analysis

ABSTRACT

Over the years, genetic studies have identified multiple genetic risk variants associated with neurodegenerative disorders and helped reveal new biological pathways and genes of interest. However, genetic risk variants commonly reside in non-coding regions and may regulate distant genes rather than the nearest gene, as well as a gene's interaction partners in biological networks. Systems biology and functional genomics approaches provide the framework to unravel the functional significance of genetic risk variants in disease. In this review, we summarize the genetic and transcriptomic studies of Alzheimer's disease and related tauopathies and focus on the advantages of performing systems-level analyses to interrogate the biological pathways underlying neurodegeneration. Finally, we highlight new avenues of multi-omics analysis with single-cell approaches, which provide unparalleled opportunities to systematically explore cellular heterogeneity, and present an example of how to integrate publicly available single-cell datasets. Systems-level analysis has illuminated the function of many disease risk genes, but much work remains to study tauopathies and to understand spatiotemporal gene expression changes of specific cell types.

1. Introduction

Alzheimer's disease (AD) is one of the most prevalent neurodegenerative disorders and the most common form of dementia, marked by both amyloid beta (A β) and tau aggregates (A β plaques and neurofibrillary tangles, respectively). Other neurodegenerative disorders (termed "tauopathies") also present pathological tau aggregates but have varied clinical manifestations, ranging from motor-related symptoms, similar to Parkinson's disease (PD), to behavioral deficits. For AD and several tauopathies, we still do not have a clear understanding of

their etiologies; however, we find shared neurological changes, such as extensive neuronal loss and glial dysfunction, suggesting that there may be shared underlying biological mechanisms across the different disorders.

Technological advancements in sequencing have provided an opportunity to uncover novel insights into the molecular bases of diseases, such as tauopathies. Genome-wide association studies (GWAS) empowered us to identify disease-associated genetic risk variants, revealing new genes of interest; however, many of these variants reside in non-coding regions, making it difficult to interpret their functional

Abbreviations: A β , Amyloid beta; AD, Alzheimer's disease; ALS, Amyotrophic Lateral Sclerosis; AMP-AD, Accelerating Medicines Partnership-Alzheimer's disease; ARM, Activated response microglia; ATAC-seq, Assay for transposase accessible chromatin with sequencing; BICCN, BRAIN Initiative Cell Census Network; CM, Co-expression module; CBD, Corticobasal degeneration; DAA, Disease-associated astrocyte; DAM, Disease-associated microglia; EC, Entorhinal Cortex; EOFAD, Early-onset familial Alzheimer's disease; FACS, Fluorescence-activated cell sorting; FTD, Frontotemporal dementia; GWAS, Genome-wide association study; GWAX, Genome-wide association study by proxy; iPSC, Induced pluripotent stem cell; IGAP, International Genomics of Alzheimer's Project; IQR, Interquartile range; LOAD, Late-onset Alzheimer's disease; MAP, Rush Memory and Aging Project; MARS-seq, Massively parallel single-cell RNA-seq; NFT, Neurofibrillary tangle; PD, Parkinson's disease; PFC, Prefrontal cortex; PPA, Primary progressive aphasia; PSP, Progressive supranuclear palsy; RNA-seq, RNA-sequencing; ROS, Religious orders study; scRNA-seq, Single-cell RNA-seq; scATAC-seq, Single-cell ATAC-seq; scVI, Single-cell variational inference; SFG, Superior frontal gyrus; snRNA-seq, Single-nucleus RNA-seq; SPLiT-seq, Split-pool ligation-based transcriptome sequencing; QTL, Quantitative trait locus; UMAP, Uniform manifold approximation and projection; WGCNA, Weighted gene co-expression network analysis.

* Corresponding author at: 3224 Biological Sciences III, Irvine, CA 92697, USA.

E-mail address: vswarup@uci.edu (V. Swarup).

¹ Co-first authors.

<https://doi.org/10.1016/j.nbd.2021.105530>

Received 14 May 2021; Received in revised form 30 August 2021; Accepted 7 October 2021

Available online 9 October 2021

0969-9961/© 2021 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

roles. Therefore, many studies have attempted to ascertain the functional significance of GWAS hits through other sequencing techniques, such as RNA-sequencing (RNA-seq) (Fig. 1). In this review, we summarize recent findings from genetic and transcriptomic studies of AD and related tauopathies. In the future, meta-analysis of high-resolution sequencing datasets will allow us to uncover disease biology with much greater sensitivity than its constituent data sources, and as a proof of principle, we performed an integrative analysis of five published single-nucleus RNA-seq (snRNA-seq) studies of human AD totaling to over 300,000 nuclei. In addition, we highlight the advantages of integrating multiple data modalities and multiple model systems and performing systems-level analyses to understand the cascade of dysregulation initiated by a genetic risk variant to yield a disease phenotype.

2. Breakthroughs and barriers in genome-wide association studies of AD and related tauopathies

GWAS in European-descended populations have identified several genetic risk factors linked to AD and related tauopathies (Table 1). Out of these conditions, most of the work in this area has focused on uncovering the heritable risk factors underlying AD. Early research into the genetics of AD largely focused on familial mutations in *APP* and *PSEN1/2*, which account for a minority of AD cases, specifically early-onset familial AD (EOFAD). The majority of AD cases though are late-onset AD (LOAD) and do not appear to have a single causal gene ('sporadic'). However, twin studies estimated that LOAD has a heritability between 60% to 80% (Gatz et al., 2006). *APOE* genotype was the first genetic risk factor identified for LOAD (Corder et al., 1993; Farrer et al., 1997; Saunders et al., 1993), accounting for about 35% of the genetic variance (Naj et al., 2011), where homozygotes for the $\epsilon 4$ allele exhibit 10- to 12-fold increased risk in comparison to those with the homozygous $\epsilon 3$ allele. On the other hand, the $\epsilon 2$ allele confers protection against AD with a 40% reduction in risk (Corder et al., 1994). *APOE* remained the only identified genetic risk factor for LOAD until the late 2000s with the first GWAS of AD, which revealed *CLU* as a novel AD risk gene (Harold et al., 2009; Lambert et al., 2009).

Large sample sizes are necessary in population genetic studies to discover disease associated risk signals. Therefore, large consortia of many different research groups have undertaken this challenge to unravel the genetic architecture of AD. For example, the International Genomics of Alzheimer's Project (IGAP) performed large-scale GWAS meta-analyses using AD patients and cognitively normal controls (Kunkle et al., 2019; Lambert et al., 2013) and has been at the forefront of AD genetics. In 2013, IGAP performed a two-stage meta-analysis of four GWAS datasets (Alzheimer's Disease Genetic Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium, the European Alzheimer's Disease Initiative, the Genetic and Environmental Risk in Alzheimer's Disease) comprising 74,046 total individuals, finding a total of 20 significant loci, 11 of which were novel at the time (Lambert et al., 2013). Earlier studies had only identified the strongest risk signals, including variants at the *APOE*, *BINI*, *CLU*, and *PICALM* loci. The most recent GWAS meta-analysis of AD from IGAP, using a larger discovery cohort than their previous study ($n = 21,982$ cases, $n = 41,944$ control, $n = 94,437$ total in all stages), confirmed 20 previously identified loci and identified five novel loci (*ADAM10*, *IQCK*, *WVVOX*, *ACE*, *ADAMTS1*) for a total of 25 significantly associated loci (Kunkle et al., 2019). From AD GWAS, we now recognize microglia as a critical cell population in AD pathophysiology; many of the identified AD risk genes (*MS46A*, *CD33*, *ABCA7*, *HLA-DRB1*, *APOE*, *TREM2*, *SPI1*, for example) are related to the immune response and/or are expressed by microglia. One of the largest hurdles blocking advancement in GWAS of AD and other complex diseases is collecting a large number of cases, and more nationwide genetics efforts such as the UK BioBank may be required for the next generation of GWAS.

While there are several AD GWAS, there are much fewer GWAS of

related tauopathies, which may be attributable in part to the relatively lower prevalence of these disorders, and these studies identified much fewer genetic variants compared to AD. Progressive supranuclear palsy (PSP), for example, is a tauopathy associated with movement impairments. Höglinger et al. (2011) performed a two-stage GWAS (stage 1: $n = 1114$ autopsy-confirmed PSP, $n = 3247$ controls; stage 2: $n = 1051$ clinically diagnosed PSP, $n = 3560$ controls) and identified three novel risk loci (*STX6*, *EIF2AK3*, and *MOBP*), in addition to *MAPT*, previously implicated in PSP risk. A GWAS meta-analysis of PSP found five loci reaching genome wide significance (*MAPT*, *MOBP*, *STX6*, *RUNX2*, and *SLCO1A2*) in a cohort of 1646 cases and 10,662 controls, which included those from Höglinger et al. (2011) (Chen et al., 2018). On the other hand, for corticobasal degeneration (CBD), which also presents motor-related symptoms, Kouri et al. (2015) examined 152 autopsy-confirmed CBD cases and 3311 controls, finding risk variants at *Inc-KIF13B-1*, *SOS1*, and *MOBP*, for example. Notably, both PSP and CBD have genetic risk variants at *MAPT*, consistent with a critical role of tau in these disorders, and at *MOBP*, suggesting oligodendrocyte dysfunction is involved in both disorders.

Frontotemporal dementia (FTD) is another common form of dementia, preferentially affecting the frontal and temporal lobes but is both clinically and pathologically heterogeneous. Pathologically, FTD is characterized by inclusions of TDP-43, FUS, or tau, and clinical subtypes include behavioral variant FTD, semantic variant primary progressive aphasia (PPA), nonfluent agrammatic variant PPA, and FTD with motor neuron disease (Gorno-Tempini et al., 2011; Neary et al., 1998; Rascofsky et al., 2011). We discuss FTD in this review due to the subset of cases with tau pathology and similar clinical symptoms with AD. Prior to GWAS, we had only identified mutations in the *MAPT*, *GRN*, *C9orf72*, *VCP*, and *CHMP2B* genes in familial FTD cases (Baker et al., 2006; Cruts et al., 2006; DeJesus-Hernandez et al., 2011; Hutton et al., 1998; Rohrer and Rosen, 2013; Skibinski et al., 2005; Watts et al., 2004; van der Zee et al., 2013), but as in AD, a large proportion of cases are sporadic. The largest GWAS of sporadic FTD to date included 3526 cases and 9402 controls in total to test for genetic association among the four clinical subtypes of FTD and performed a meta-analysis of all the samples together, finding significant signals at the *HLA* and *BTNL2* loci (Ferrari et al., 2014). Notably, *HLA* is also an AD risk gene and implicates the immune response in both AD and FTD. However, in this study, FTD cases were determined primarily by clinical diagnoses, with only 3% pathologically confirmed, restricting analysis of FTD risk by its pathological subtypes, and similar to PSP and CBD, the sample numbers are substantially smaller than AD GWAS, thus impeding the identification of additional risk variants.

Due to the large heritable component of many complex traits, the analytical capability of GWAS can be expanded by including individuals with a family history of disease, in addition to primary cases. In 2017, Liu, Elrich, and Pickrell introduced an approach called genome-wide association study by proxy (GWAX), where they replaced cases with first-degree family members to analyze 12 common disorders using UK Biobank data, and the associations found in these GWAX were comparable to those of GWAS with primary cases (Liu et al., 2017). This may be especially useful for rarer disorders, such as PSP and CBD, where it may be more difficult to obtain large sample numbers with primary cases alone. Following this initial study, several groups have used this approach for AD GWAS meta-analysis including cases, controls, and familial proxy cases (Jansen et al., 2019; Marioni et al., 2018; Schwartzentruber et al., 2021). The AD-by-proxy study from Jansen et al. (2019) identified 29 significant risk loci, including novel signals at *ADAMTS4*, *KAT8*, *CLNK*, and several other loci, while Schwartzentruber et al. (2021) identified a total of 37 loci, altogether expanding the catalog of known AD-associated genetic signals. Given its track record in many disorders, GWAX should be successful in further uncovering the genetic architecture of related tauopathies.

Researchers have also been particularly interested in investigating shared genetic risk across neurodegenerative disorders. Ferrari et al.

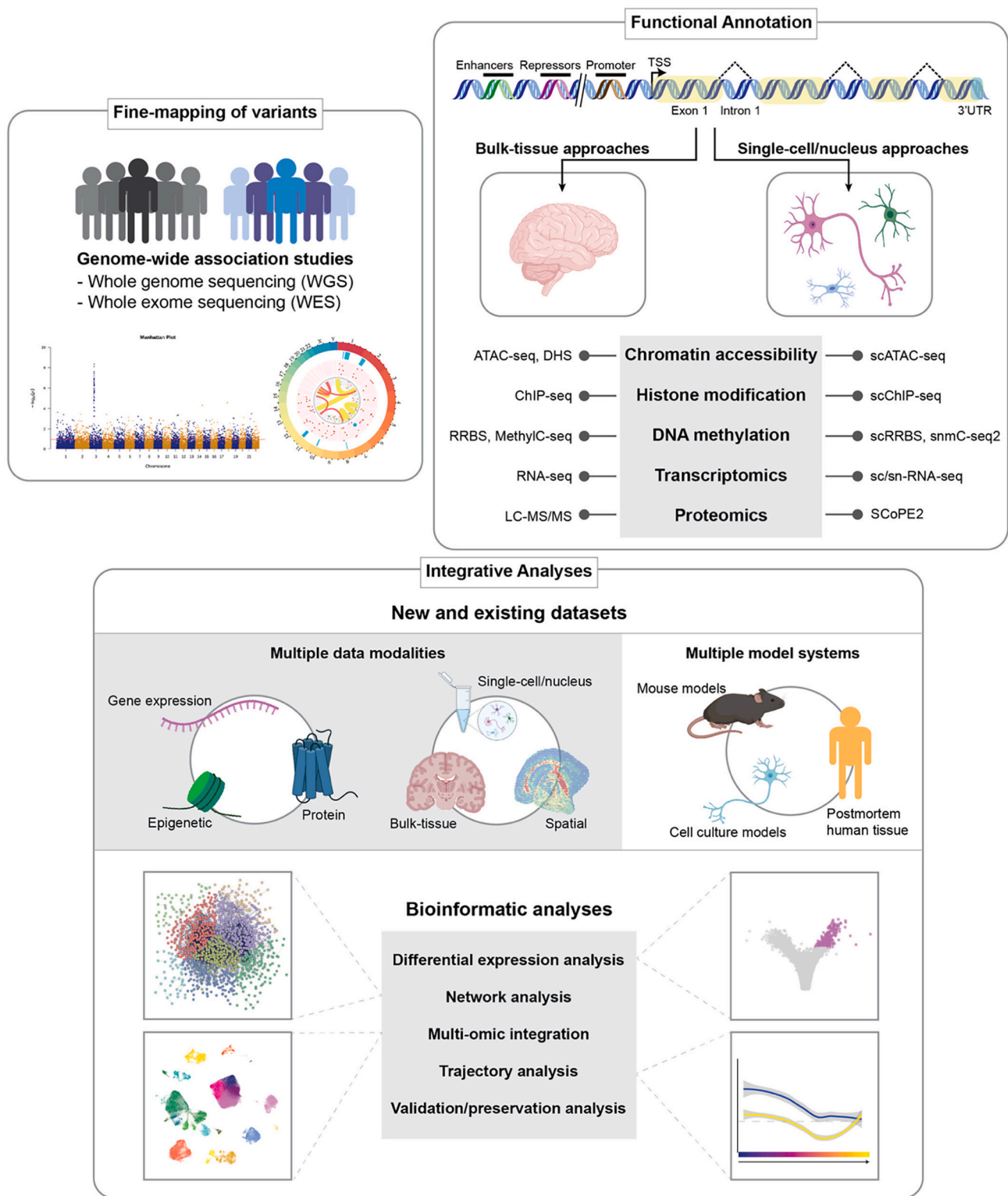


Fig. 1. Sequencing techniques and analysis for understanding genetic variants.

Schematic diagram for functional analysis of GWAS variants. Genome wide association studies use genotyping chips, whole-genome sequencing, and whole-exome sequencing to reveal genomic loci that are enriched for disease-risk signals, and subsequent genetic fine mapping analysis homes in on likely causal disease risk variants at a given GWAS locus. High-throughput genomics approaches at bulk-tissue or single-cell/nucleus resolution in relevant tissues provide functional annotations at disease risk loci. There are many different approaches to characterize different aspects of the genome; for example ChIP-seq reveals genomic sequences that are bound by a transcription factor or have a histone modification of interest, while ATAC-seq identifies sequences in accessible chromatin regions. RNA-seq provides abundance information for all RNA transcripts in each sample. Integrative bioinformatics approaches combining a variety of data modalities and model systems can elucidate the mechanisms underlying genetic risk variants. Multi-scale information from the epigenome, transcriptome, and proteome can be leveraged from different model systems in downstream bioinformatics analysis to prioritize disease-relevant signatures.

Fig. 1 was created with BioRender.

Table 1
Summary of GWAS of AD and related tauopathies.

Title (Author, Year)	Trait	N cases	N controls	total loci	novel loci	nearest genes to risk loci	Ancestry included
Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease (Lambert et al., 2013)	Alzheimer's Disease	Stage1: 17,008 Stage 2: 8572	Stage 1: 37,154 Stage 2: 11,312	20	11	<i>APOE, CR1, BIN1, CD2AP, EPHA1, CLU, MS4A6A, PICALM, ABCA7, CD33, HLA-DRB5/HLA-DRB1, PTK2B, SORL1, SLC24A5/RIN3, DSG2, INPP5D, MEF2C, NME8, ZCWPW1, ZELF1, FERMT2, CASS4</i>	European
Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ε4, and the Risk of Late-Onset Alzheimer's disease in African Americans (Reitz et al., 2013)	Alzheimer's Disease	1968	3928	NR*	1	<i>APOE, ELMO1, SOX13, ANCA7, GRIN3B, HMA1, CR1, BIN1, PICALM, CLU, EPHA1, MS4A, CD2AP, CD33</i>	African American
Transethnic genome-wide scan identifies novel Alzheimer's disease loci (Jun et al., 2017)	Alzheimer's Disease	Stage1: 15,5798 Stage 2: 5813	Stage 1: 17,690 Stage 2: 20,474	20	9	<i>PFDN1/HBEGF, USP6NL/ECHDC3, BZRAP1-AS1, CR1, BIN1, PTK2B, CLU, MS4A4A, PICALM, ABCA7, NFIC, TPBG, HBEGF, NME8, CASS4</i>	European, African-American, Japanese, Israeli-Arabs
GWAS on family history of Alzheimer's disease (Marioni et al., 2018)	Family history of Alzheimer's Disease	27,696 Maternal AD; 14,338 Paternal AD	260,980	26	6	<i>CR1, BIN1, PILRA, PICALM, ZNF232, APOE, TOMM40, SCIMP, SPPL2A, IL-34, PLCG2, ADAM10, BCKDK/KAT8, ACE, TREML2, VKORC1, BZRAP1, CNLK, ADAMTS4, CCDC6</i>	European
Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing (Kunkle et al., 2019)	Alzheimer's Disease	Stage 1: 21,982 Stage 2: 8362 Stage 3A: 4930	Stage 1: 41,944 Stage 2: 10,483 Stage 3A: 6736	25	5	<i>CR1, BIN1, INPP5D, HLA-DRB1, TREM2, CD2AP, NYAP1, EPHA1, PTK2B, CLU, SPI1, MS4A2, PICALM, SORL1, FERMT2, SLC24A4, ABCA7, APOE, CASS4, ECHDC3, ACE</i>	European
Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk (Jansen et al., 2019)	Alzheimer's Disease, AD-by-proxy	Phase 1: 24,087 Phase 2 (proxy): 47,793 Replication: 6593	Phase 1: 55,058 Phase 2 (proxy): 328,320 Replication: 174,289	29	9	<i>ADAMTS4, CR1, BIN1, INPP5D, HESX1, CLNK, HLA-DRB1, TREM2, CD2AP, ZCWPW1, EPHA1, CNTNAP2, CLU/PTK2B, ECHDC3, MS4A6A, PICALM, SORL1, SLC24A4, ADAM10, APH1B, KAT8, SCIMP, ABI3, ALPK2, ABCA7, AC074212.3, CD33, CASS4</i>	European
Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes (Schwartzentruber et al., 2021)	Alzheimer's Disease, AD-by-proxy	53,042	355,900	37	4	<i>ADAMTS4, CR1, SPRED2, NCK2, BIN1, TMEM163, INPP5D, CLNK, HLA-DRB1, UNC5CL, CD2AP, SPATA48, PILRA, EPHA1, CLU, ECHDC3, CCDC6, TSPAN14, SPI1, MS4A6E, PICALM, SORL1, FERMT2, SLC24A4, SPPL2A, ADAM10, APH1B, KAT8, PLCG2, SCIMP, TSPPOAP1, ACE, ABCA7, APOE, CD33, CASS4, ADAMTS1</i>	European
Novel Alzheimer's disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis (Kunkle et al., 2021)	Alzheimer's Disease	2748	5222	20	11	<i>TRANK1, FABP2, LARP1B, TSRM, ARAP1, STARD10, SPHK1, SERPINB13, SIPA1L2, EDEM1, ALCAM, ACER3, PIK3C2G, IGF1R, RBFOX1, VRK3, WDR70</i>	African American
Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk (Shigemizu et al., 2021)	Alzheimer's Disease	3692	4074	9	1	<i>APOE, SORL1, FAM47E, PAPOLG, RAB3C, BANK1, LINC01867, LINC00899, LOC101928561, APOE, PICALM, BIN1, CLU, CR1, MS4A4A, SORL1, MADD, HLA-DRA, CD2AP, OR2B2, EPHA1, ADAMTS1, SLC24A4, LACTB2, ELL, FERMT2, ZCWPW1, TSPPOAP1</i>	Japanese, European (trans-ethnic meta-analysis)
Frontotemporal dementia and its subtypes: a genome-wide association study (Ferrari et al., 2014)	Frontotemporal dementia	Stage 1: 3526	Stage 1: 9402	NR*	2	<i>RAB38, CTSC, HLA-DRA/HLA-DRB5, BTNL2, MAPT, C9orf72/MOB3B, TMEM106B, TOMM40/APOE</i>	European
Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy (Höglinger et al., 2011)	Progressive supranuclear palsy	Stage 1: 1114 Stage 2: 1051	Stage 1: 3247 Stage 2: 3560	3	3	<i>STX6, EIF2AK3, MOBP, MAPT</i>	European
Joint genome-wide association study of progressive supranuclear palsy identifies novel	Progressive supranuclear palsy	1646	10,662	5	2	<i>MAPT, MOBP, STX6, EIF2AK3, SEMA4D, DDX27, SPI1, RUNX2, DUSP10, WDR63, MIR4423,</i>	European

(continued on next page)

Table 1 (continued)

Title (Author, Year)	Trait	N cases	N controls	total loci	novel loci	nearest genes to risk loci	Ancestry included
susceptibility loci and genetic correlation to neurodegenerative diseases (Chen et al., 2018)						<i>SLCO1A2, ASAP1, AMHR2, CEP57, RPS6K1</i>	
Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy (Kouri et al., 2015)	Corticobasal degeneration	Discovery: 152 Replication: 67	Discovery: 3311 Replication: 439	4	3	<i>MAPT, Inc-KIF13B-1, SOS1, MOBP</i>	European

(2017) examined the genetic overlap between sporadic FTD and AD or PD and found one variant at the *APOE* locus jointly associated with FTD and AD. Interestingly, they also identified variants at *HLA* and *MAPT* loci shared between FTD and PD but not AD, indicating a distinct genetic architecture of AD (Ferrari et al., 2017). Chen et al. (2018) also investigated the genetic correlation between PSP and AD or FTD but did not find a significant correlation for either disorder. Instead, they found significant overlaps between PSP and Amyotrophic Lateral Sclerosis (ALS) or PD genetic risk (Chen et al., 2018). Another study identified genetic overlap between ALS and FTD as well (Karch et al., 2018). ALS and PD are typically associated with TDP-43 and alpha-synuclein inclusions, respectively, and thus these genetic studies imply that there may be shared molecular mechanisms driving neurodegeneration independent of protein pathology and suggest new avenues for cross-disorder therapeutic development.

GWAS have illuminated many genetic risk signals for AD and related tauopathies thus strengthening our understanding of neurodegeneration, but there are some key limitations of these studies that need to be addressed in the future. To date, the majority of GWAS in AD and other disorders have only profiled European-descendant populations, which is problematic because different ethnic groups have distinct genetic risk profiles. For example, the *APOE* $\epsilon 4$ allele is associated with an increased risk of AD in Japanese populations but a diminished risk in African American populations in comparison to European populations (Graff-Radford et al., 2002; Miyashita et al., 2013). There is a clear need for the generation and analysis of more large-scale GWAS datasets of non-European populations to further our understanding of polygenic diseases like AD.

Additionally, while GWAS are useful for pinpointing variants that are likely causal for disease, this analysis is merely a starting point for the broader task of understanding how genetic variation disrupts the normal function of specific tissues and cell types that thereby contributes to the manifestation of a disease phenotype. Therefore, studies have leveraged expression quantitative trait loci and epigenetic datasets to try to functionally characterize genetic risk variants; however, these datasets are not necessarily from human brain tissue samples, potentially impeding accurate predictions. While deciphering the biology of genetic risk variants in humans remains a daunting task, biotechnological advances and systems genomics may lead us closer towards this goal.

3. Surveying the diseased brain using co-expression network analysis

To understand the genetic basis of neurodegenerative disorders, we require a comprehensive picture portraying the cascade of interactions between disease-risk signals and cis-regulatory elements, genes, and proteins. Integration of genetic studies with high-throughput functional genomics may lead us towards this goal. The next-generation genomics revolution has paved the way for neuroscientists to study genome-wide molecular signatures of the human brain, such as gene expression, at an unprecedented speed. Multiple studies have performed RNA-sequencing (RNA-seq) or microarray-based gene expression analysis on AD post-mortem human brain tissue samples, and to a lesser extent in FTD and PSP, and identified upregulated and downregulated genes in the

respective disorders (Allen et al., 2018; Berchtold et al., 2013; Jiang et al., 2018; Rexach et al., 2020; Wang et al., 2016; Webster et al., 2009; Zhang et al., 2013). For example, it is now well established that AD GWAS gene *BIN1* is upregulated in AD (Chapuis et al., 2013; Holler et al., 2014; Karch et al., 2012; Martiskainen et al., 2015). Additionally, several studies have identified disease-associated microRNAs (miRNAs) and splicing variants, improving our understanding of alterations in gene regulation with disease (Lau et al., 2013; Raj et al., 2018; Swarup et al., 2019).

However, these studies primarily rely on a gene-level statistical analysis, such as differential expression testing, providing little knowledge about relationships between different genes. Therefore, it is important to also employ a more holistic analysis to achieve a systems-level understanding of gene expression and a more comprehensive view of brain circuits and functions in health and disease. Co-expression network analysis is a powerful systems biology approach that can identify multi-scale gene expression patterns across brain region, cell-types, cortical layers, and neural circuits (Oldham et al., 2008; Parikshak et al., 2013; Wang et al., 2016; Zhang et al., 2013). This hierarchical organization provides a platform for integrating different types of data, including multi-omics approaches, imaging, and behavioral studies. Network analysis complements differential expression analysis by providing structure to the underlying data, resembling the intricate organization inherent to biological systems. In addition, network analysis goes beyond pathway-based analysis approaches, which are inherently biased towards known biological perspectives and thus hamper discovery-centric data-driven research.

Although different analytical strategies have been used for co-expression network analysis (Chiu and Talhouk, 2018; Gaiteri et al., 2015; Langfelder and Horvath, 2008; Song and Zhang, 2015; Tesson et al., 2010), the general underlying principle is to organize gene (or protein) expression data into groups of modules or communities, comprising nodes (genes or gene-products) and edges, which can represent physical interactions (like protein-protein interactions) or more commonly correlation or mutual information. The connectivity of the edges can be leveraged to understand intra-network (also known as intra-modular) connectivity between different nodes and highlight the most connected genes ("hub" genes) in a network. Several *in-silico* and functional genomic analyses have shown that co-expression modules are functional biological units (Carter et al., 2013; Chandran et al., 2016; Geschwind and Konopka, 2009; Mitra et al., 2013), and studying the group properties of these modules in health and disease have helped us understand the molecular underpinnings of biological systems.

Co-expression network analysis performed on both microarray and RNA-sequencing datasets of human AD identified several biological pathways dysregulated with disease, such as those related to synaptic transmission, mitochondrial function, and immune response (Forabosco et al., 2013; Miller et al., 2008; Morabito et al., 2020; Mostafavi et al., 2018; Wang et al., 2016; Zhang et al., 2013). Although initial studies had small sample sizes of under or around 100 individuals, many of the identified biological pathways were also identified in studies of hundreds, if not a thousand individuals, indicating that network analysis identifies robust, disease-relevant gene networks. In addition, many AD-correlated modules are conserved across different brain regions

(Forabosco et al., 2013; Morabito et al., 2020; Wang et al., 2016). Multiregional analyses also have indicated regional specificity of AD gene expression changes that reflect the pathological progression of AD; gene expression changes in regions with early tau deposition, such as the hippocampus, are more pronounced than those in regions later affected in the disease (Morabito et al., 2020; Wang et al., 2016). Altogether, these co-expression studies highlighted several genes as key regulators of these biological pathways, laying the groundwork for future studies in AD, and notably helped to clarify the role of microglial genes, such as *TYROBP* and *TREM2*, in AD genetic risk. For example, network analysis revealed *TYROBP* regulates an immune-related module correlated with AD (Zhang et al., 2013).

Network analysis has also highlighted the importance of oligodendrocytes in AD pathophysiology. Oligodendrocytes have been relatively understudied in AD compared to astrocytes and microglia, despite multiple studies revealing AD-related white matter changes (Ihara et al., 2010; Nasrabady et al., 2018). *PSEN1* mutations are known to result in

EOFAD (Cacace et al., 2016), and co-expression network analysis identified *PSEN1* is co-expressed with myelin-related genes (Miller et al., 2008; Zhang et al., 2013). Additionally, myelination-related modules are dysregulated in PSP, indicating cross-disorder oligodendrocyte dysfunction (Allen et al., 2018).

Many of these studies, however, have analyzed data from a single brain tissue repository, and differences in tissue processing can affect downstream results. In our previous study, we constructed co-expression modules from data generated by the Accelerating Medicines Partnership-Alzheimer’s Disease (AMP-AD) consortium, consisting of more than 1000 human AD and control samples from different brain regions across multiple brain banks (Morabito et al., 2020). We used consensus weighted gene co-expression network analysis (cWGCNA), a meta-analytical approach, to identify robust disease-specific co-expressed modules, which we identified as neuronal or non-neuronal (Fig. 2a-b). These gene co-expression changes were preserved in the normal human aging brain but were altered only with the progression of AD (Fig. 2c-d).

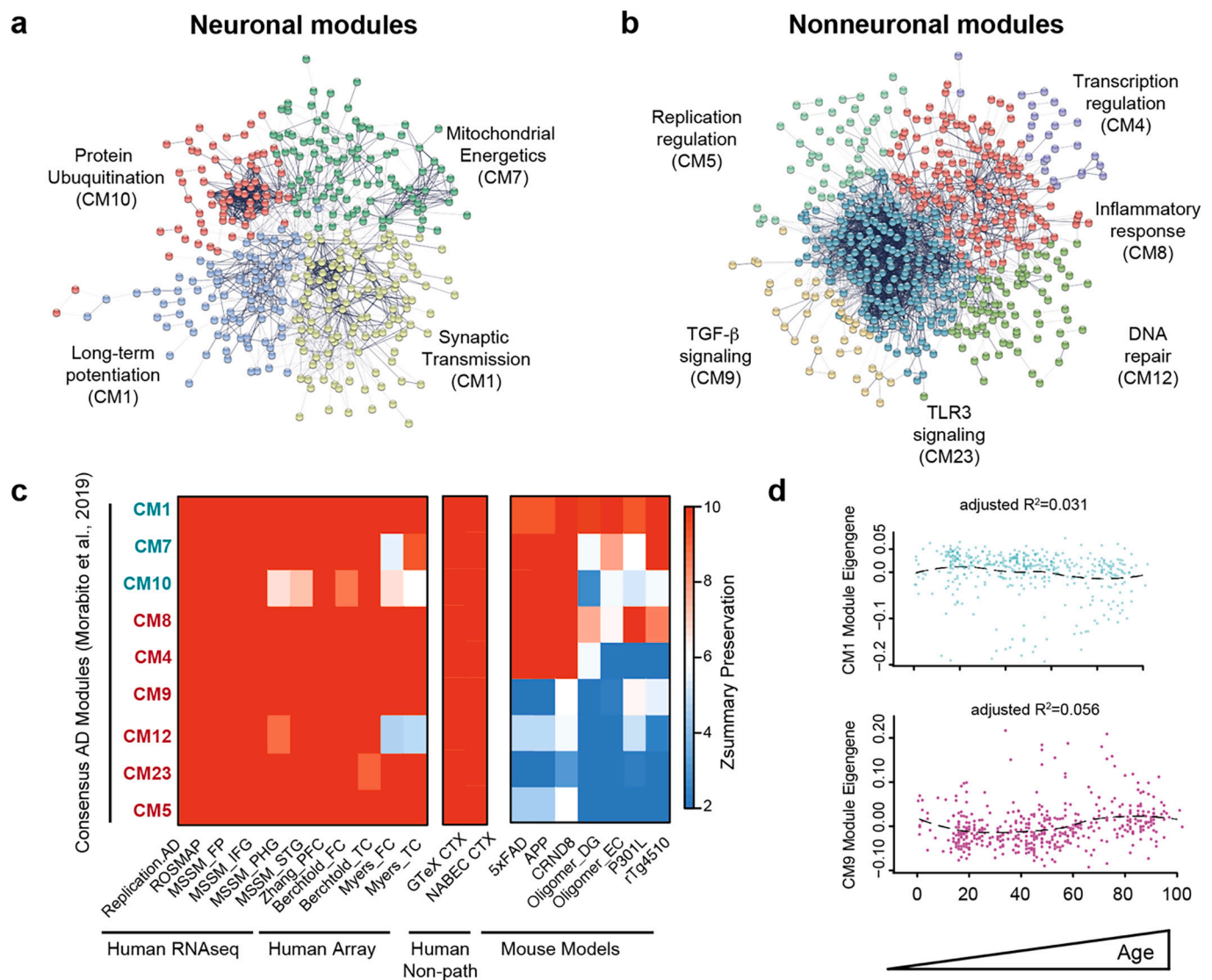


Fig. 2. AD consensus gene co-expression modules.

a, b, Neuron-specific (**a**) and glia-specific (**b**) co-expression modules (CM) significantly correlated with AD diagnosis, annotated with their related biological functions. In these networks, each node represents a gene, colored by co-expression module (gene group), and each edge represents a link between a pair of genes based on co-expression and protein-protein interaction evidence. This visualization provides a general overview of the module size and relationships between genes in the modules. **c,** AD consensus co-expression modules are robust. Heatmap showing modules are preserved across human AD RNA-seq and microarray datasets, as well as datasets from human non-pathological controls (non-path) and mouse models. **d,** AD co-expression changes are specific to disease. Scatterplot show module eigengenes CM1 (top) and CM9 (bottom) are not correlated with age.

We also found that the co-expression modules were highly preserved across several additional human AD microarray datasets, and their trajectories changed similarly in these datasets, demonstrating that the gene co-expression changes identified via network analysis are robust, despite technical differences. Additionally, we deeply characterized our consensus co-expression modules by integrating various orthogonal datasets including GWAS, DNA methylation, histone acetylation (H3K9ac), and expression and methylation quantitative loci (eQTLs and mQTLs), to provide a comprehensive picture of the genetic and transcriptomic alterations associated with AD. We identified a microglial module (CM8) that was significantly enriched in AD GWAS genes, supporting the key role of microglia in AD genetic risk, and our analysis revealed transcription factors, such as SPI1 and IRF8, as potential regulators of this module.

Moreover, co-expression network analysis approaches provide a platform for integrating data from different model systems of disease. This provides us the potential to discover disease-associated, evolutionary conserved changes. Notably, we found that some of the glial modules, such as those related to TGF β and TLR3 signaling, were not preserved in mouse models, such as the 5XFAD, a transgenic mouse model of AD expressing five human AD gene mutations (*APP* and *PSEN1*) that result in extensive amyloid deposition, and rTg4510, a tau mouse model with the human MAPT P301L mutation, associated with FTD (Fig. 2c), possibly highlighting the lack of translatability from mouse models to human clinical trials (Drummond and Wisniewski, 2017; Onos et al., 2016; Vitek et al., 2020). Neuronal modules though were generally preserved across both tau and A β mouse models (Fig. 2c). A recent study also constructed co-expression modules from AMP-AD datasets but utilized 5 different co-expression network analysis approaches (MEGENA (Song and Zhang, 2015), WINA (Wang et al., 2016), rWGCNA (Parikshak et al., 2016), speakeasy (Gaiteri et al., 2015), and their own metanetwork), identifying similar modules across the different methods (Wan et al., 2020). Wan et al. (2020) integrated multiple AD and other neurodegeneration-associated mouse models at different stages of disease progression to identify conserved gene co-expression changes. Importantly, they found overlaps between human AD co-expression modules and differentially expressed genes in mouse models of Huntington's disease and spinocerebellar ataxia, disorders not typically associated with AD, suggesting a generalized transcriptomic signature of neurodegeneration, not specific to AD. Additionally, our previous work (Swarup et al., 2019) used mouse models of AD and FTD, induced pluripotent stem cell (iPSC) derived neurons, and human postmortem samples to identify two co-expression modules a) neurodegeneration-associated synaptic (NAS) module and b) neurodegeneration-associated inflammatory (NAI) module, which were validated in a dozen other mouse models of disease. We identified a miRNA (miR-203) as a novel regulator of the NAS module.

Proteomic studies of AD and PSP have leveraged network analysis in order to define the disease-associated proteome and relate genetic risk to protein expression (Johnson et al., 2020; Seyfried et al., 2017; Swarup et al., 2020). These studies revealed that there are disease proteomic networks that are distinct from those found at the transcriptome level. In our previous work (Swarup et al., 2020), we integrated five different proteomic datasets using consensus WGCNA to robustly identify disease-associated proteomic modules, and we found enrichment of AD and PSP GWAS hits in separate modules, indicating that their etiologies are distinct, despite the presence of tau pathology in both disorders. Integration of different data modalities, such as transcriptomic, proteomic, and genetic data, provides a more comprehensive understanding of the molecular changes associated with disease.

4. Single-cell genomics has uncovered vast cellular dysregulation in Alzheimer's disease

Recent advances in sequencing methods have enabled us to profile single cells dissociated from whole tissues (Cao et al., 2017;

Hashimshony et al., 2012; Jaitin et al., 2014; Klein et al., 2015; Macosko et al., 2015; Ramsköld et al., 2012; Rosenberg et al., 2018; Zheng et al., 2017). In "bulk-tissue" sequencing studies of whole tissue samples, the signal of interest is averaged across all the cells present in the sample. Thus, bulk sequencing techniques are limited in their capacity to faithfully characterize the biology of the brain, which is composed of many different types of neurons and glia with fine-grained regional specificity. Single-cell sequencing approaches employ nucleotide barcoding strategies that allow us to bioinformatically separate all the sequencing reads to their cell of origin. Where a bulk-tissue RNA-seq experiment of one sample yields only one data point describing each gene, a single-cell RNA-seq (scRNA-seq) experiment of the same sample could yield hundreds up to tens of thousands of data points describing each cell, depending on the specific single-cell approach. This presents us with new opportunities for characterizing how specific cell types respond to disease, especially in the context of disease risk signals, as well as introducing many new challenges in terms of best practices for data analysis (Lähnemann et al., 2020; Luecken and Theis, 2019).

There are several different approaches for single-cell sequencing, each with their own advantages and disadvantages in terms of throughput and data quality (Ziegenhain et al., 2017). Over a decade ago, the first single-cell sequencing study (Tang et al., 2009) was published, where a single mouse blastomere was manually picked and libraries were prepared for mRNA-seq. Fortunately, modern approaches do not require manual cell picking, rather they use specialized microfluidics devices or combinatorial nucleotide barcoding schemes to prepare single cells or single nuclei for sequencing. To date, scRNA-seq is the most widely used single-cell genomics method, but other techniques, such as single-cell assay for transposase-accessible chromatin with high-throughput sequencing (scATAC-seq), are gaining traction as they are commercialized. A new method called split-pool ligation-based transcriptome sequencing (SPLiT-seq) (Rosenberg et al., 2018) is an attractive cost-effective approach for neurobiologists that are interested in performing their own single-cell sequencing since it does not require a microfluidic device.

Several studies have used sc- or snRNA-seq to characterize the cellular heterogeneity of the AD brain in human patients and mouse models of disease (Grubman et al., 2019; Habib et al., 2020; Keren-Shaul et al., 2017; Lau et al., 2020; Leng et al., 2021; Mathys et al., 2019; Zhou et al., 2020). In general, while each of these studies have used different experimental designs and statistical approaches, all of them have contributed to our understanding of dysregulation of specific cell types and cell subpopulations in AD. The first transcriptomic study of the human AD brain at cellular resolution used the 10 \times Genomics platform to perform snRNA-seq in 48 prefrontal cortex (PFC) samples from the Religious Order Study (ROS) and the Rush Memory and Aging Project (MAP) (Mathys et al., 2019). This study resolved gene expression profiles of the major cell types of the PFC and identified differentially expressed genes between individuals with AD pathology versus those with no pathology for each of these cell types. Furthermore, this study identified several cell subpopulations that are overrepresented in certain pathological and biological conditions, such as Braak stage, cognitive status, or sex. For example, one subtype of excitatory neurons termed Ex6 (*STMN1*⁺, *PARK7*⁺, *FBXO2*⁺) was more common to male samples in Braak stages one and two, whereas excitatory cluster Ex4 (*LINGO1*⁺, *RASGEF1B*⁺, *SLC26A3*⁺) was enriched in female samples from Braak stage five. Additional sex-specific cell type responses to AD were found in this study, such as a distinct transcriptional activation of oligodendrocytes in males and a clear downregulation of neuronal gene expression in females with respect to pathological variables.

More snRNA-seq studies of AD have followed the initial study from Mathys et al. (2019), one of which being a study that profiled the entorhinal cortex (EC) in six AD patients and six cognitively normal controls, where the authors specifically investigated cell-type-specific gene expression signatures of AD genetic risk genes (Grubman et al., 2019). For each of the major cell types, they identified significant differences in

gene expression of AD risk genes across different cell subpopulations, highlighting *BIN1* in a specific astrocyte population and *CLU* in an oligodendrocyte progenitor population, for example. Furthermore, selective vulnerability is a well-known feature of neurodegeneration, where different neuronal populations exhibit a range of susceptibility to disease. Another recent study used snRNA-seq to characterize the EC and the superior frontal gyrus (SFG) in postmortem tissue from male donors with varying Braak stages to investigate selective vulnerability of neurons in AD, revealing *RORB* as a marker gene for vulnerable excitatory neurons and noting a distinct lack of vulnerability among inhibitory neuron populations (Leng et al., 2021). A key lesson from these different snRNA-seq studies of AD is that gene expression is widely dysregulated throughout disease progression across cell types and brain regions, but different subpopulations of cells are more responsive or resilient to disease.

Prior to any of these human snRNA-seq studies, one of the first single-cell studies in an AD mouse model (Keren-Shaul et al., 2017) used massively parallel single-cell RNA-seq (MARS-seq) to compare gene expression in brain immune cells between wild type and 5XFAD mice. Their analysis uncovered a subpopulation of microglia enriched in aged 5XFAD mice, which they termed “Disease Associated Microglia” (DAMs), and they revealed the genes that are dynamically expressed throughout the change from homeostatic to disease-associated cell states. The DAM state is marked by an increased expression of AD-associated risk genes, such as *ApoE* and *Trem2*, and a decreased expression of homeostatic markers, such as *Cx3cr1* and *P2ry12*. Other studies identified a similar cell state distinct from homeostatic microglia and enriched in diseased samples, which have been termed “Activated Response Microglia” (ARMs) (Frigerio et al., 2019; Sierksma et al., 2020).

Additionally, another group performed a snRNA-seq study of the PFC using samples from human AD patients, including those with the common variant of *TREM2* (*TREM2*-CV) and those with the AD risk variant *TREM2*-R62H, and *Trem2*-deficient 5XFAD mice (Zhou et al., 2020). They found that DAM genes, such as *ApoE* and *Csf1*, are upregulated in 5XFAD compared to *Trem2*-deficient 5XFAD mice, while homeostatic genes including *Cx3cr1* were downregulated, providing support for the *Trem2*-dependent DAM hypothesis. In contrast, by comparing gene expression in microglia between human controls and AD (*TREM2*-CV) samples, only a handful of DAM genes, including *TREM2*, *CD68*, and *APOE*, were upregulated in human AD. This study also identified several key differences between human AD and the 5XFAD mouse model in astrocytes and oligodendrocytes.

Our group recently performed a cross-species snRNA-seq analysis using three human AD datasets and one 5XFAD dataset, and we found one microglia population that was enriched for the DAM gene signature (Morabito et al., 2021). By performing differential gene expression analysis comparing mouse and human microglia nuclei in this cluster, we found a distinct depletion of many DAM genes in human microglia. Despite the deep characterization of DAMs and DAM-like cells in mouse models, it is not clear whether these *Trem2*-dependent A β responsive microglia identified in mice directly correspond to disease responsive microglia in human AD brains. Furthermore, another study (Habib et al., 2020) identified a population of A β -localized disease responsive astrocytes, termed “Disease Associated Astrocytes” (DAAs), and in our group’s single-cell analysis we found evidence that the DAA gene signature is present in human AD samples (Morabito et al., 2021).

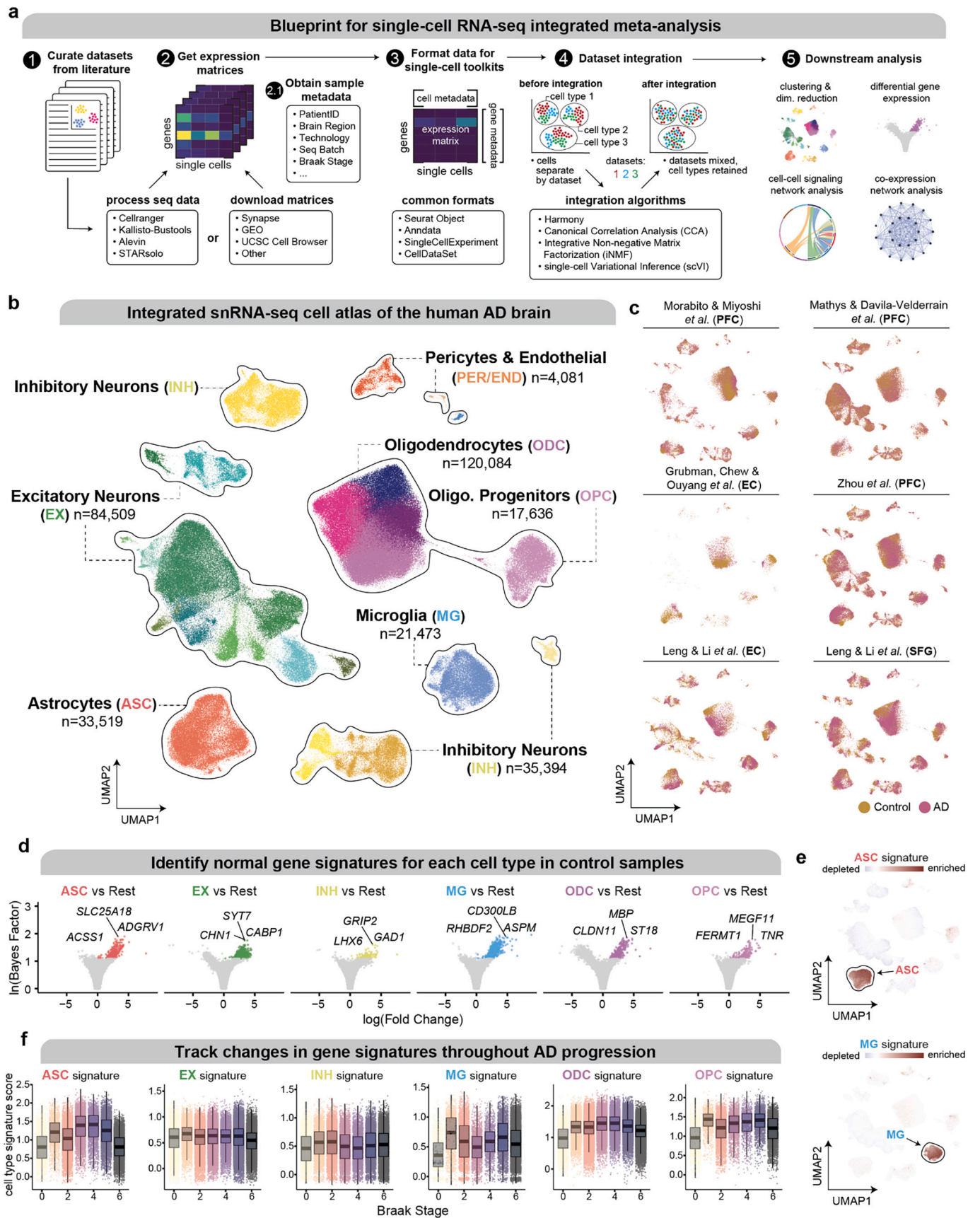
Much work has been done to characterize the transcriptome of AD at cellular resolution; however, many key questions remain unanswered. For instance, while these studies have demonstrated cell type specific gene expression of AD risk genes, and while there is evidence for the role of some of these risk genes in the glial response to A β , it is not clear how these risk factors disrupt the gene regulatory networks in the healthy brain to contribute towards the molecular phenotypes of disease. Epigenetic data can aid in clarifying the gene regulatory mechanisms in disease. For example, Nott et al. (2019) generated cell-type specific

chromatin conformation and histone modification data from fluorescence-activated cell sorting (FACS) of human brain nuclei and identified a microglia-specific *BIN1* enhancer mapping to an AD risk variant. However, in this study, they were unable to distinguish epigenetic changes unique to specific cell subpopulations (Nott et al., 2019), and it is now possible to generate single-cell epigenetic data (without FACS). Similar to single-cell transcriptomic studies, single-cell chromatin accessibility studies have identified heterogeneous subsets of cells in the human brain (Corces et al., 2020; Lake et al., 2017; Morabito et al., 2021). Corces et al. (2020) predicted regulatory targets of AD risk variants in loci such as *PICALM* and *BIN1* in specific cell types, but the sequencing data originated from cognitively normal samples. Our recent study profiled single-cell chromatin accessibility from both late-stage pathology AD samples and cognitively normal controls, and we constructed cell type specific cis-gene regulatory networks (Morabito et al., 2021). We compared the chromatin accessibility landscape and predicted cis-regulatory interactions between cells from AD and control samples at selected GWAS loci, including *APOE* and *BIN1*, and found cell-type-specific differences flanking the likely causal SNPs, thus highlighting potential interactions between these non-coding SNPs and neighboring regulatory elements, which may underlie disease. Furthermore, we integrated snATAC-seq and snRNA-seq data from the same samples to identify candidate cis-regulatory elements and their potential target genes by cell type, highlighting the strength of integrating single-cell transcriptomic and epigenetic datasets to interrogate the molecular mechanisms underlying neurodegeneration.

5. Integrating single-cell datasets for meta-analysis

Single-cell sequencing, however, is still rather expensive in comparison to whole tissue approaches, hindering our ability to associate cellular signatures to pathological features, given the small sample size in the present single-cell studies. For example, large-scale studies of bulk AD transcriptomics have jointly analyzed over 1000 human samples, while to date the Mathys & Davila-Velderrain et al. study has the largest number of samples used in a single-cell study of human AD with 48 samples. However, it is possible to greatly increase the analytical power of an individual study by leveraging other published datasets in the form of a single-cell meta-analysis. Recent breakthroughs in single-cell integration techniques have made it possible to perform large-scale meta-analysis of single-cell data from multiple studies, species, or -omics modalities (Argelaguet et al., 2021). As an example of what can be done with these integration approaches, the BRAIN Initiative Cell Census Network (BICCN) has constructed a multimodal cross-species cell atlas of the motor cortex that combines data from snRNA-seq, scRNA-seq, single-cell DNA methylation (snmC-seq), snATAC-seq, spatial transcriptomics (MERFISH), and other technologies (Adkins et al., 2020).

Here we present a blueprint for constructing an integrated single-cell atlas of AD, and we note that this workflow is applicable to different meta-analysis approaches, which could include multiple species, data modalities, or different disease conditions (Fig. 3a). As an example of this integrative analysis, we curated six snRNA-seq datasets of human AD encompassing three brain regions and 124 samples ($n = 65$ AD, $n = 59$ control, see Supplementary Methods). There are several different integration approaches that would be adequate for this task (Hie et al., 2019; Korsunsky et al., 2019; Stuart et al., 2019; Welch et al., 2019), but here we performed integration using single-cell Variational Inference (scVI), a deep neural network that models the biological information underlying measured gene expression values (Lopez et al., 2018). scVI is an ideal approach for creating a cell atlas of AD because it can be updated to incorporate new datasets after the initial training step. After training the scVI model, we performed dimensionality reduction using uniform manifold approximation and projection (UMAP), a standard tool for visualizing snRNA-seq data in two or three dimensions (Becht et al., 2019), and we grouped cells into clusters using Leiden community detection (Traag et al., 2019). Using this integrative framework, we



(caption on next page)

Fig. 3. snRNA-seq meta-analysis of Alzheimer's disease.

a. Schematic diagram for integrating multiple single-cell transcriptomics datasets. The general principles here can also be applied to other single-cell datasets. **b.** Cell populations from six AD snRNA-seq datasets (Grubman et al., 2019; Leng et al., 2021; Mathys et al., 2019; Morabito et al., 2021; Zhou et al., 2020) integrated with single-cell Variational Inference (scVI) are visualized with uniform manifold approximation and projection (UMAP). Each dot represents a single nucleus and is colored by cell subpopulation (cell cluster, Leiden cluster assignment). Major cell lineages are circled and annotated. **c.** UMAP plots as in (b), colored based on AD diagnosis and split by study of origin, show how the distributions of the cell populations may differ based on brain region, disease diagnosis, or other technical variables between the different datasets. Ideally, cell populations should be mixed between the datasets, not originating from only one dataset. **d.** Gene expression signatures for each major cell type by identifying cell type marker genes. Volcano plots summarize iterative "one cell type versus rest" differential gene expression tests for each cell type in control brains only, showing log-scaled fold-change and Bayes factors for each gene. Genes with moderate significance are colored, using a threshold of $\ln(\text{Bayes Factor}) > 1.1$, and the top three genes by significance are annotated. **e.** Cell type marker genes, identified through differential expression analysis, define transcriptionally distinct cell populations by major cell type. UMAP plots as in (b), colored by single cell gene signatures based on the top 20 marker genes from the tests in (d) for astrocytes (top) and microglia (bottom). **f.** Box and whisker plots show how the distribution of cell type signatures as in (e) change with increasing Braak stage. Distributions for these cell type signatures are visualized only for cells in the corresponding cell types, e.g. the astrocyte signature is plotted in astrocytes only. The range of the box corresponds to the inter-quartile range (IQR), the whiskers extend to the lowest or highest values at most 1.5 times the IQR, and the median is shown within the box. The underlying data points are plotted behind the box and whisker plots.

recovered the same major cell lineages from the original studies (Fig. 3b), and we found that the data was well integrated across different studies and brain regions (Fig. 3c). As an example of what could be done with this integrated dataset, we sought to track how the gene signature of each major cell type changes throughout the progression of AD. Using nuclei from control samples only (Braak stages 0, 1, 2), we used the trained scVI model to perform differential gene expression tests iteratively comparing each major cell type to all other cells, thus identifying a robust set of cell-type-specific marker genes (Fig. 3d). We then took the top 20 marker genes, sorted by statistical significance in each cluster, and computed "gene module scores" for each cell in the dataset, which represent a common cell type signature present in non-AD brains (Fig. 3e). Finally, we plotted the distributions of these cell-type gene signatures throughout different Braak stages using cells in their respective cell types, thus showing from a broad perspective that gene expression profiles of each cell type are dynamic throughout the progression of disease (Fig. 3f). In this integrated dataset, we found both astrocyte and oligodendrocyte gene signatures increase from no pathology to mid-stage pathology before decreasing with advanced pathology, possibly reflecting an initial glial response to pathology that declines when the insult becomes too great in the late stages of disease. On the other hand, both excitatory and inhibitory neuron gene signatures are relatively stable across Braak stages. These gene signatures can also be dissected by brain region to determine region specificity of these gene expression changes, in addition to a subpopulation-level analysis.

This example analysis demonstrates how future single-cell transcriptomics studies of AD can take advantage of the previous work that has been done in this field through integration techniques to further study aspects of cellular dysfunction of AD that have not been explored in previous studies. In addition to these published datasets, there are several exciting pre-prints in this area, namely one study that developed a custom protocol to isolate NFT-bearing neurons for single-cell sequencing (Otero-Garcia et al., 2020), another that specifically profiled patients with monogenic AD (Marinero et al., 2020), and a study that profiled the brain's vasculature in AD (Yang et al., 2021). Tauopathies have not been investigated using single-cell genomics, but this integrative analysis approach could be used to identify shared and distinct patterns of cellular dysregulation across disorders. In sum, this analysis demonstrates the feasibility of using many different data sources to perform a unified single-cell meta-analysis.

6. Future directions

Our knowledge of the function of genetic risk factors for AD and related tauopathies has been greatly accelerated by innovation in high-throughput genomics methods. While there are many emerging -omics approaches that could be used to discover new insights into neurodegeneration, many of these techniques are only used by a small number of labs, since they are not commercially available. On the other hand, while some of these cutting-edge techniques, such as single-cell

transcriptomics, have been used to study AD, there are still no published studies of tauopathies at cellular resolution. Furthermore, there are very few high-resolution epigenomics studies of AD or related tauopathies. Over the coming years, we can expect to learn more about the relationship between genetic risk and molecular dysfunction in tauopathies by using -omics and systems-level data analysis.

Transcriptomic dysregulation of the brain in disease can be better understood with spatial information, which is lost in sc- or snRNA-seq. A recent study used spatial transcriptomics to profile AD gene expression changes with spatial context in human patients and a mouse model, identifying gene expression patterns that are associated with the spatial organization of amyloid deposits (Chen et al., 2020). However, many spatial transcriptomics methods do not profile gene expression at cellular resolution, which can make it difficult to disentangle which cell type is driving a certain signal at any given location. High-resolution spatial approaches such as RNA seqFISH+ (Eng et al., 2019) and integrative bioinformatics approaches can now resolve transcriptional information in single cells (Kleshchevnikov et al., 2020), and a preliminary study has identified gene expression profiles of the dendrites and soma of individual neurons (Wang et al., 2020). Additionally, RNA splicing information is lost in many single-cell approaches that use Illumina sequencing, which does not sufficiently capture full length RNA transcripts. New methods for long-read single-cell sequencing can unravel transcript isoforms that are dysregulated with disease in a cell-type-specific manner (Gupta et al., 2018; Joglekar et al., 2021; Rebboah et al., 2021).

High-resolution epigenomic sequencing methods are rapidly emerging, but to date very few have been used to study the diseased brain. Our group's previous work (Morabito et al., 2021) was the first to use both snATAC-seq and snRNA-seq in the same samples to characterize disease-associated cellular dysregulation in AD, and one other study analyzed AD and PD associated risk loci using snATAC-seq in control brains (Corces et al., 2020). Cell-type-specific histone modifications and DNA-protein interactions can be measured using single-cell CUT&Tag (Bartosovic et al., 2021). There are also methods to study DNA methylation (Luo et al., 2018) and 3D chromatin interactions in single cells (Ramani et al., 2017). Furthermore, there are now several sequencing approaches that profile more than one -omic modality in the same cell, such as Paired Tag for profiling the transcriptome and histone modifications (Zhu et al., 2021); SNARE-seq (Chen et al., 2019) and 10× Genomics Multiome to profile the transcriptome and chromatin accessibility; and TEA-seq for profiling the transcriptome, chromatin accessibility, and epitopes (Swanson et al., 2021). In light of these new developments in biotechnology and bioinformatics, there are many opportunities to elucidate the underlying mechanisms of neurodegeneration in tauopathies, to systematically uncover how genetic risk factors play a role in the dysregulation of specific cell types and develop new therapeutics targeting novel regulators of disease.

- Swarup, V., Hinz, F.I., Rexach, J.E., Noguchi, K., Toyoshiba, H., Oda, A., Hirai, K., Sarkar, A., Seyfried, N.T., Cheng, C., Haggarty, S.J., Ferrari, R., Rohrer, J.D., Ramasamy, A., Hardy, J., Hernandez, D.G., Nalls, M.A., Singleton, A.B., Kwok, J.B.J., Dobson-Stone, C., Brooks, W.S., Schofield, P.R., Halliday, G.M., Hodges, J.R., Piguet, O., Bartley, L., Thompson, E., Haan, E., Hernández, I., Ruiz, A., Boada, M., Borroni, B., Padovani, A., Cairns, N.J., Cruchaga, C., Binetti, G., Ghidoni, R., Benussi, L., Forloni, G., Albani, D., Galimberti, D., Fenoglio, C., Serpente, M., Scarpini, E., Clarimón, J., Lleó, A., Blesa, R., Waldó, M.L., Nilsson, K., Nilsson, C., Mackenzie, I.R.A., Hsiung, G.-Y.R., Mann, D.M.A., Grafman, J., Morris, C.M., Attems, J., Griffiths, T.D., McKeith, I.G., Thomas, A.J., Jaros, E., Pietrini, P., Huey, E. D., Wassermann, E.M., Tierney, M.C., Baborie, A., Pastor, P., Ortega-Cubero, S., Razuquin, C., Alonso, E., Pernecky, R., Diehl-Schmid, J., Alexopoulos, P., Kurz, A., Rainero, I., Rubino, E., Pinessi, L., Rogava, E., George-Hyslop, P.St., Rossi, G., Tagliavini, F., Giaccone, G., Rowe, J.B., Schlachetzki, J.C.M., Uphill, J., Collinge, J., Mead, S., Danek, A., Deerlin, V.M.V., Grossman, M., Trojanowski, J.Q., Pickering-Brown, S., Momeni, P., van der Zee, J., Cruts, M., Broeckhoven, C.V., Cappa, S.F., Leber, I., Brice, A., Hannequin, D., Galfier, V., Vercelletto, M., Nacmias, B., Sorbi, S., Bagnoli, S., Piaceri, I., Nielsen, J.E., Hjerfjord, L.E., Riemenschneider, M., Mayhaus, M., Gasparoni, G., Pichler, S., Ibach, B., Rossor, M.N., Fox, N.C., Warren, J. D., Spillantini, M.G., Morris, H.R., Rizzu, P., Heutink, P., Snowden, J.S., Rollinson, S., Gerhard, A., Richardson, A., Bruni, A.C., Maletta, R., Frangipane, F., Cupidi, C., Bernardi, L., Anfossi, M., Gallo, M., Conidi, M.E., Smirne, N., Rademakers, R., Baker, M., Dickson, D.W., Graff-Radford, N.R., Petersen, R.C., Knopman, D., Josephs, K.A., Boeve, B.F., Parisi, J.E., Miller, B.L., Karydas, A.M., Rosen, H., Seeley, W.W., van Swieten, J.C., Dopper, E.G.P., Seelaar, H., Pijnenburg, Y.A.L., Scheltens, P., Logroscino, G., Capozzo, R., Novelli, V., Poca, A.A., Franceschi, M., Postiglione, A., Milan, G., Sorrentino, P., Kristiansen, M., Chiang, H.-H., Graff, C., Pasquier, F., Rollin, A., Deramecourt, V., Lebouvier, T., Ferrucci, L., Kapogiannis, D., Grossman, M., Deerlin, V.M.V., Trojanowski, J.Q., Lah, J.J., Levey, A.I., Kondou, S., Geschwind, D.H., 2019. Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat Med* 25, 152–164. <https://doi.org/10.1038/s41591-018-0223-3>.
- Swarup, V., Chang, T.S., Duong, D.M., Dammer, E.B., Dai, J., Lah, J.J., Johnson, E.C.B., Seyfried, N.T., Levey, A.I., Geschwind, D.H., 2020. Identification of conserved proteomic networks in neurodegenerative dementia. *Cell Reports* 31, 107,807. <https://doi.org/10.1016/j.celrep.2020.107807>.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., Lao, K., Surani, M.A., 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6, 377–382. <https://doi.org/10.1038/nmeth.1315>.
- Tesson, B.M., Breitling, R., Jansen, R.C., 2010. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *Bmc Bioinformatics* 11, 497. <https://doi.org/10.1186/1471-2105-11-497>.
- Traag, V.A., Waltman, L., van Eck, N.J., 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 9, 5233. <https://doi.org/10.1038/s41598-019-41,695-z>.
- van der Zee, J., Gijssels, I., Dillen, L., Langenhove, T.V., Theuns, J., Engelborghs, S., Philtjens, S., Vandenbulcke, M., Sleegers, K., Sieben, A., Bäumer, V., Maes, G., Corsmit, E., Borroni, B., Padovani, A., Archetti, S., Pernecky, R., Diehl-Schmid, J., de Mendonça, A., Miltenberger-Miltenyi, G., Pereira, S., Pimentel, J., Nacmias, B., Bagnoli, S., Sorbi, S., Graff, C., Chiang, H.-H., Westerlund, M., Sanchez-Valle, R., Llado, A., Gelpi, E., Santana, I., Almeida, M.R., Santiago, B., Frisoni, G., Zanetti, O., Bonvicini, C., Synofzik, M., Maetzler, W., Hagen, J.M.V., Schöls, L., Heneka, M.T., Jessen, F., Matej, R., Parobkova, E., Kovacs, G.G., Ströbel, T., Sarafov, S., Tournev, I., Jordanova, A., Danek, A., Arzberger, T., Fabrizi, G.M., Testi, S., Salmon, E., Santens, P., Martin, J.-J., Cras, P., Vandenbergh, R., De Deyn, Peter Paul, Cruts, M., van der Broeckhoven, C.V., Zee, J., Gijssels, I., Dillen, L., Langenhove, T.V., Theuns, J., Philtjens, S., Sleegers, K., Bäumer, V., Maes, G., Corsmit, E., Cruts, M., Broeckhoven, C.V., van der Zee, J., Gijssels, I., Dillen, L., Langenhove, T.V., Philtjens, S., Theuns, J., Sleegers, K., Bäumer, V., Maes, G., Cruts, M., Broeckhoven, C.V., Engelborghs, S., De Deyn, Peter P., Cras, P., Engelborghs, S., Deyn, De Peter P., Vandenbulcke, M., Vandenbulcke, M., Borroni, B., Padovani, A., Archetti, S., Pernecky, R., Diehl-Schmid, J., Synofzik, M., Maetzler, W., Hagen, J.M.V., Schöls, L., Synofzik, M., Maetzler, W., Hagen, J.M.V., Schöls, L., Heneka, M.T., Jessen, F., Ramirez, A., Kurzwelly, D., Sachtleben, C., Mairer, W., de Mendonça, A., Miltenberger-Miltenyi, G., Pereira, S., Firmo, C., Pimentel, J., Sanchez-Valle, R., Llado, A., Antonell, A., Molinuevo, J., Gelpi, E., Graff, C., Chiang, H.-H., Westerlund, M., Graff, C., Ståhlbom, A.K., Thonberg, H., Nennesmo, I., Börjesson-Hanson, A., Nacmias, B., Bagnoli, S., Sorbi, S., Bessi, V., Piaceri, I., Santana, I., Santiago, B., Santana, I., Ribeiro, M.H., Almeida, M.R., Oliveira, C., Massano, J., Garret, C., Pires, P., Frisoni, G., Zanetti, O., Bonvicini, C., Sarafov, S., Tournev, I., Jordanova, A., Tournev, I., Kovacs, G.G., Ströbel, T., Heneka, M.T., Jessen, F., Ramirez, A., Kurzwelly, D., Sachtleben, C., Mairer, W., Jessen, F., Matej, R., Parobkova, E., Danel, A., Arzberger, T., Fabrizi, G.M., Testi, S., Ferrari, S., Cavallaro, T., Salmon, E., Santens, P., Cras, P., Consortium, E.E.-O.D., 2013. A pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability, and intermediate repeats. *Hum Mutat* 34, 363–373. <https://doi.org/10.1002/humu.22244>.
- Vitek, M.P., Aratjio, J.A., Fossel, M., Greenberg, B.D., Howell, G.R., Rizzo, S.J.S., Seyfried, N.T., Tenner, A.J., Territo, P.R., Windisch, M., Bain, L.J., Ross, A., Carrillo, M.C., Lamb, B.T., Edelmayer, R.M., 2020. Translational animal models for Alzheimer's disease: an Alzheimer's association business consortium think tank. *Alzheimer's Dementia Transl Res Clin Interventions* 6, e12114. <https://doi.org/10.1002/trc2.12114>.
- Wan, Y.-W., Al-Ouran, R., Mangleburg, C.G., Perumal, T.M., Lee, T.V., Allison, K., Swarup, V., Funk, C.C., Gaiteri, C., Allen, M., Wang, M., Neuner, S.M., Kaczorowski, C.C., Phillip, V.M., Howell, G.R., Martini-Stoica, H., Zheng, H., Mei, H., Zhong, X., Kim, J.W., Dawson, V.L., Dawson, T.M., Pao, P.-C., Tsai, L.-H., Haure-Mirande, J.-V., Ehrlich, M.E., Chakrabarty, P., Levites, Y., Wang, X., Dammer, E.B., Srivastava, G., Mukherjee, S., Sieberts, S.K., Omberg, L., Dang, K.D., Eddy, J.A., Snyder, P., Chae, Y., Amberkar, S., Wei, W., Hide, W., Preuss, C., Ergun, A., Ebert, P. J., Airey, D.C., Mostafavi, S., Yu, L., Klein, H.-U., Consortium, A.M.P.-A.D., Carter, G. W., Collier, D.A., Golde, T.E., Levey, A.I., Bennett, D.A., Estrada, K., Townsend, T.M., Zhang, B., Schadt, E., Jager, P.L.D., Price, N.D., Ertekin-Taner, N., Liu, Z., Shulman, J.M., Mangravite, L.M., Logsdon, B.A., 2020. Meta-analysis of the Alzheimer's disease human brain transcriptome and functional dissection in mouse models. *Cell Reports* 32, 107,908. <https://doi.org/10.1016/j.celrep.2020.107908>.
- Wang, M., Roussos, P., McKenzie, A., Zhou, X., Kajiwara, Y., Brennand, K.J., Luca, G.C. D., Cray, J.F., Casaccia, P., Buxbaum, J.D., Ehrlich, M., Gandy, S., Goate, A., Katsel, P., Schadt, E., Haroutunian, V., Zhang, B., 2016. Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med* 8, 104. <https://doi.org/10.1186/s13073-016-0355-3>.
- Wang, G., Ang, C.-E., Fan, J., Wang, A., Moffitt, J.R., Zhuang, X., 2020. Spatial organization of the transcriptome in individual neurons. *Biorxiv* 2020 (12), 07.414060. <https://doi.org/10.1101/2020.12.07.414060>.
- Watts, G.D.J., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P., Kimonis, V.E., 2004. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet* 36, 377–381. <https://doi.org/10.1038/ng1332>.
- Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., McCorquodale, D.S., Cuello, C., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Josphira, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J.V., Group, N.-N., Heward, C.B., Reiman, E.M., Stephan, D., Hardy, J., Myers, A.J., 2009. Genetic control of human brain transcript expression in Alzheimer's disease. *Am J Hum Genetics* 84, 445–458. <https://doi.org/10.1016/j.ajhg.2009.03.011>.
- Welch, J.D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., Macosko, E.Z., 2019. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* 177. <https://doi.org/10.1016/j.cell.2019.05.006>, 1873–1887.e17.
- Yang, A.C., Vest, R.T., Kern, F., Lee, D.P., Maat, C.A., Losada, P.M., Chen, M.B., Agam, M., Schaum, N., Khoury, N., Calcuttawala, K., Pálovics, R., Shin, A., Wang, E. Y., Luo, J., Gate, D., Siegenthaler, J.A., McNERney, M.W., Keller, A., Wyss-Coray, T., 2021. A human brain vascular atlas reveals diverse cell mediators of Alzheimer's disease risk. *Biorxiv* 2021 (04), 26.441262. <https://doi.org/10.1101/2021.04.26.441262>.
- Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezchnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., Fluder, E., Clurman, B., Melquist, S., Narayanan, M., Suver, C., Shah, H., Mahajan, M., Gillis, T., Mysore, J., MacDonald, M.E., Lamb, J.R., Bennett, D.A., Molony, C., Stone, D.J., Gudnason, V., Myers, A.J., Schadt, E.E., Neumann, H., Zhu, J., Emilsson, V., 2013. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 153, 707–720. <https://doi.org/10.1016/j.cell.2013.03.030>.
- Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., Gregory, M.T., Shuga, J., Montesclaros, L., Underwood, J.G., Masquelier, D.A., Nishimura, S.Y., Schnall-Levin, M., Wyatt, P.W., Hindson, C.M., Bharadwaj, R., Wong, A., Ness, K.D., Beppu, L.W., Deeg, H.J., McFarland, C., Loeb, K.R., Valente, W.J., Ericson, N.G., Stevens, E.A., Radich, J.P., Mikkelsen, T.S., Hindson, B.J., Bielas, J.H., 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8, 14,049. <https://doi.org/10.1038/ncomms14049>.
- Zhou, Y., Song, W.M., Andhey, P.S., Swain, A., Levy, T., Miller, K.R., Poliani, P.L., Cominelli, M., Grover, S., Gilfillan, S., Cella, M., Ulland, T.K., Zaitsev, K., Miyashita, A., Ikeuchi, T., Sainouchi, M., Kakita, A., Bennett, D.A., Schneider, J.A., Nichols, M.R., Beausoleil, S.A., Ulrich, J.D., Holtzman, D.M., Artyomov, M.N., Colonna, M., 2020. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med* 26, 131–142. <https://doi.org/10.1038/s41591-019-0695-9>.
- Zhu, C., Zhang, Y., Li, Y.E., Lucero, J., Behrens, M.M., Ren, B., 2021. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat Methods* 18, 283–292. <https://doi.org/10.1038/s41592-021-01060-3>.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., Enard, W., 2017. Comparative analysis of single-cell RNA sequencing methods. *Mol Cell* 65. <https://doi.org/10.1016/j.molcel.2017.01.023>, 631–643.e4.