



## Review paper

## Integrative multi-omics and systems bioinformatics in translational neuroscience: A data mining perspective

Lance M. O'Connor<sup>a</sup>, Blake A. O'Connor<sup>b</sup>, Su Bin Lim<sup>c</sup>, Jialiu Zeng<sup>d</sup>, Chih Hung Lo<sup>d,\*</sup><sup>a</sup> College of Biological Sciences, University of Minnesota, Minneapolis, MN, 55455, USA<sup>b</sup> School of Pharmacy, University of Wisconsin, Madison, WI, 53705, USA<sup>c</sup> Department of Biochemistry and Molecular Biology, Ajou University School of Medicine, Suwon, 16499, South Korea<sup>d</sup> Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, 308232, Singapore

## ARTICLE INFO

## Article history:

Received 29 November 2022

Received in revised form

20 June 2023

Accepted 25 June 2023

Available online 30 June 2023

## Keywords:

Multi-omics integration

Systems bioinformatics

Data mining

Human brain profile reconstruction

Translational neuroscience

## ABSTRACT

Bioinformatic analysis of large and complex omics datasets has become increasingly useful in modern day biology by providing a great depth of information, with its application to neuroscience termed neuroinformatics. Data mining of omics datasets has enabled the generation of new hypotheses based on differentially regulated biological molecules associated with disease mechanisms, which can be tested experimentally for improved diagnostic and therapeutic targeting of neurodegenerative diseases. Importantly, integrating multi-omics data using a systems bioinformatics approach will advance the understanding of the layered and interactive network of biological regulation that exchanges systemic knowledge to facilitate the development of a comprehensive human brain profile. In this review, we first summarize data mining studies utilizing datasets from the individual type of omics analysis, including epigenetics/epigenomics, transcriptomics, proteomics, metabolomics, lipidomics, and spatial omics, pertaining to Alzheimer's disease, Parkinson's disease, and multiple sclerosis. We then discuss multi-omics integration approaches, including independent biological integration and unsupervised integration methods, for more intuitive and informative interpretation of the biological data obtained across different omics layers. We further assess studies that integrate multi-omics in data mining which provide convoluted biological insights and offer proof-of-concept proposition towards systems bioinformatics in the reconstruction of brain networks. Finally, we recommend a combination of high dimensional bioinformatics analysis with experimental validation to achieve translational neuroscience applications including biomarker discovery, therapeutic development, and elucidation of disease mechanisms. We conclude by providing future perspectives and opportunities in applying integrative multi-omics and systems bioinformatics to achieve precision phenotyping of neurodegenerative diseases and towards personalized medicine.

© 2023 The Authors. Published by Elsevier B.V. on behalf of Xi'an Jiaotong University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The beginning of bioinformatics can be dated back to more than five decades ago, witnessing the parallel advances of computer science and experimental biology [1], including the advent of next-generation sequencing and omics technologies [2]. The omics technologies and analyses have been widely applied in neuroscience studies, ranging from the detection of alterations in genes (epigenetics/epigenomics), mRNA (transcriptomics), proteins (proteomics), and metabolites (metabolomics/lipidomics) at the molecular scale in the brain, their localization in different

anatomical regions (spatial omics) to construct brain atlases, and the study of dynamics in biological processes by monitoring changes in individual cells (single-cell/single-nucleus trajectory inference) [3–8]. Importantly, these omics studies provide a great depth of information and contain datasets that can be further analyzed and interpreted to support separate experimental observations or generate new hypotheses in neuroscience research [9].

Data mining of existing omics datasets to obtain novel biological insights opens new avenues to deepen the understanding of the clinical phenotypes, neuropathological features, disease progression, and pathogenic mechanisms of neurological disorders including Alzheimer's disease (AD), Parkinson's disease (PD) and multiple sclerosis (MS) [10]. In addition, data mining overcomes the technical challenges in conducting the initial experiments such as difficulty in obtaining precious tissue samples, stringent requirements in sample

Peer review under responsibility of Xi'an Jiaotong University.

\* Corresponding author.

E-mail address: [chihhung.lo@ntu.edu.sg](mailto:chihhung.lo@ntu.edu.sg) (C.H. Lo).

preparation, and high cost, while creating opportunities for new diagnostic and therapeutic strategies targeting neurodegenerative diseases. While single omics analyses have provided valuable disease mechanistic insights, recent studies have shown that integrative multi-omics analysis can help to define the connection and relationship among the different types of omics datasets to unravel brain networks regulating transitions from health to the development of neurological diseases and to classify clinically relevant subgroups to identify potential biomarkers [11,12].

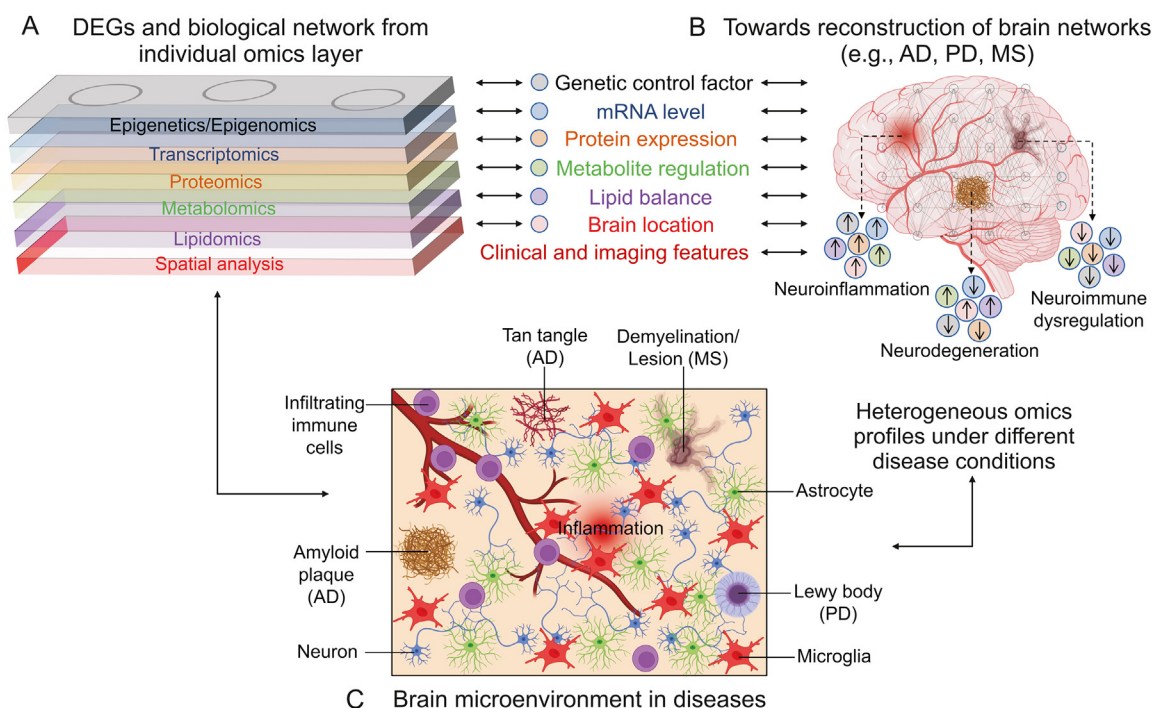
It is hence crucial to understand the neurodegenerative pathology from the perspective of individual type of omics data obtained from the human samples, followed by adopting the systems bioinformatics approach to integrate the multiscale and multi-source big data from each omics layer to allow for a holistic analysis of the complex brain system (Fig. 1A). Importantly, there is heterogeneous omics profiles under various disease conditions, such as neuroinflammation, neurodegeneration, and neuro-immune dysregulation, which can be characterized by different combinations of alterations in the omics layers (Fig. 1B). In addition, further development and optimization of multi-omics integration approaches and pipelines will potentially enable the reconstruction of comprehensive brain networks and pathological profiles reflective of the biological systems and the microenvironments under specific disease states (Fig. 1C) [5,13–16]. Furthermore, there is a need for thorough interpretation of the outcomes from data mining and their relevance to the true biological observations.

In this review, we first summarize data mining based translational neuroscience studies that performed secondary bioinformatics analysis utilizing deposited datasets from a spectrum of omics technologies including epigenetics, transcriptomics, proteomics, metabolomics, lipidomics, and spatial omics. We then discuss current methods of multi-omics integration and propose a systems

bioinformatics approach to work towards the reconstruction of brain networks and pathological profiles for increased accuracy and reliability in recapitulating the true biological systems of the brain. We further recommend combining high dimensional bioinformatics analysis with experimental validation in biomarker discovery, therapeutic development, and elucidation of disease mechanisms. We conclude by providing future perspectives and opportunities in utilizing integrative multi-omics and systems bioinformatics to achieve targeted therapies for neurodegenerative diseases and advance towards personalized medicine.

## 2. Omics spectrum-based data mining in translational neuroscience

Many factors contribute to neurodegeneration, including but not limited to, pathogenic mutations leading to mutant protein production and toxic protein aggregation [17–20], as well as altered cytokine production, and dysregulated cellular signaling pathways [21,22]. As opposed to conventional interventions aimed at specific toxic protein or aberrant receptor signaling, interventions at the gene level can more effectively reduce the negative downstream cellular pathogenic effects that arise from malfunctioning of a single protein. Gene targeting strategies can be particularly beneficial in neurodegenerative diseases that result from the toxic gain-of-function of a protein where a significant loss of its normal function does not have adverse effects on the cells [23]. Furthermore, there is a need for better understanding of the regional effects of a gene or protein in the brain as well as the time-dependent changes in their expression or functions to enable specific targeting and more effective treatments. In this section, we will summarize the individual omics-based data mining studies in translational neuroscience and their applications in guiding biomarker and therapeutic development.



**Fig. 1.** Towards reconstruction of brain networks and pathological profiles of neurodegenerative diseases by systems bioinformatics. (A) Integrated network of multi-omics data can be obtained from the individual omics layer. A combination of integrated networks with known clinical and imaging features may provide detailed molecular information towards reconstruction of neurodegenerative disease pathological profiles in the brain. (B) The presence of heterogeneous omics profiles under various disease conditions, such as neuroinflammation, neurodegeneration, and neuroimmune dysregulation, which can be characterized by different combinations of alterations in the omics layers. (C) The reconstructed brain networks could reflect the pathological profiles associated with the true biological systems and the microenvironments under specific disease states. The figure was created with [BioRender.com](https://www.biorender.com). DEGs: differentially expressed genes; AD: Alzheimer's disease; PD: Parkinson's disease; MS: multiple sclerosis.

**Table 1**

List of databases and repositories that are available for data mining of different types of omics datasets.

Databases and repositories	Data types	Refs.
DeepBlue Epigenomic Data Server	Epigenetics/Epigenomics	[31]
Epigenome-Wide Association Study (EWAS) Open Platform	Epigenetics/Epigenomics	[32]
Genomic Expression Archive (GEA)	Epigenetics/Epigenomics	[33]
Genome Wide Associated Studies (GWAS) Catalog	Epigenetics/Epigenomics	[34]
International Human Epigenome Consortium (IHEC) Portal	Epigenetics/Epigenomics	[35]
National Cell Repository for Alzheimer's Disease (NCRAD)	Epigenetics/Epigenomics	[36]
Roadmap Epigenomics	Epigenetics/Epigenomics	[37]
Gene Expression Omnibus (GEO) Database	Epigenetics/Epigenomics/Bulk RNA-seq	[38]
Mount Sinai Brain Bank	Bulk RNA-seq	[54]
ROSMAP Database	Bulk RNA-seq	[55]
ARCHS4 Database	Bulk RNA-seq/sc/snRNA-seq	[56]
DDBJ Sequence Read Archive (SRA)	Bulk RNA-seq/sc/snRNA-seq	[57]
Synapse Database	Bulk RNA-seq/sc/snRNA-seq	[58]
scREAD Database	sc/snRNA-seq	[69]
Allen Brain Map	sc/snRNA-seq	[70]
DRscDB Database	sc/snRNA-seq	[71]
Global Proteome Machine Database (GPMDB)	Proteomics	[87]
jPOSTrepo (Japan ProteOme S'Tandard Repository)	Proteomics	[88]
MassIVE	Proteomics	[89]
Proteomic Data Commons	Proteomics	[90]
ProteomeXchange	Proteomics	[91]
PRoteomics IDentifications Database (PRIDE)	Proteomics/Spatial Omics	[92]
ProteomicsDB	Proteomics/Spatial Omics	[93]
Alzheimer's Disease Neuroimaging Initiative (ADNI)	Metabolomics/Lipidomics	[110]
Cerebrospinal Fluid Metabolome Database	Metabolomics/Lipidomics	[111]
Human Metabolome Database	Metabolomics/Lipidomics	[112]
Lipid Bank	Metabolomics/Lipidomics	[117]
LipidBlast	Metabolomics/Lipidomics	[118]
Lipid MAPS	Metabolomics/Lipidomics	[119]
MetaboAge Database	Metabolomics/Lipidomics	[120]
MetaboLights	Metabolomics/Lipidomics	[121]
Metabolomics Workbench	Metabolomics/Lipidomics	[122]
MetabolomeXchange	Metabolomics/Lipidomics	[123]
Serum Metabolome Database	Metabolomics/Lipidomics	[124]
Omics Discovery Index (OmicsDI)	All Omics	[127]
Dynamic Proteomics	Spatial Omics	[165]
Giotto	Spatial Omics	[166]
Spatial TranscriptOmics DataBase (STOmicsDB)	Spatial Omics	[167]
SpatialDB	Spatial Omics	[168]

## 2.1. Epigenetics/epigenomics analysis

With recent advancements in experimental and computational tools to analyze neurodegenerative diseases, a genetic basis for these diseases has yet to be fully elucidated [24]. In this section, we will discuss the gene-centric view of understanding the role of epigenetics/epigenomics in the pathogenesis of neurodegenerative diseases. Epigenetics can be summarized as heritable changes in gene function which are not encoded by nucleotide sequences in DNA, but influence gene expression and subsequent protein expression levels without altering the DNA sequence [25,26]. These changes can be due to major epigenetic mechanisms such as DNA methylation, histone modifications, chromatin remodeling, non-coding RNA regulation, as well as environmental factors such as diet, and exposure to chemicals which are dynamic and reversible [24–27], making them viable targets for therapeutic developments [28]. Experimental studies exploring the role of epigenetic mechanisms in neurodegenerative diseases have found that DNA methylation variations may affect beta-amyloid (A $\beta$ ) and tau deposition in AD, and expression of  $\alpha$ -synuclein in PD [29,30]. While epigenetics/epigenomics datasets are deposited in various databases, including DeepBlue [31], EWAS Open Platform [32], Genomic Expression Archive (GEA) [33], Genome Wide Associated Studies (GWAS) [34], IHEC Data Portal [35], National Cell Repository for Alzheimer's disease (NCRAD) [36], Roadmap Epigenomics [37], and Gene Expression Omnibus (GEO) [38] (Table 1, Epigenetics/Epigenomics database), there are limited number of data mining studies.

This has led to the need to overcome practical impediments including developing algorithms and models for large-scale data mining of epigenetics/epigenomics datasets [39]. One data mining study utilizing DNA methylation data obtained from the GEO database implemented a supervised machine learning algorithm, including the construction of differential network related to aging acceleration and the use of Markov Chain Monte Carlo method of global sensitivity analysis, to better understand the accelerated epigenetic aging mechanisms of various neurodegenerative diseases [40]. Their results indicated that individuals with neurodegenerative diseases exhibited a significantly accelerated aging pattern. Specifically, they found that *CDCA7L* and *EFNB2* are significantly different than other genes in AD and PD, respectively. While *CDCA7L* is involved in neuronal death, and *EFNB2* is involved in apoptosis and the development of the nervous system as well as neuronal migration. Their analysis further revealed that *DUSP12* had the largest betweenness across different disease types, and that *DUSP12* may regulate the c-Jun N-terminal kinase signaling pathway by dephosphorylating its substrate, which is critical to cell differentiation, apoptosis, and other neural functions in the progression of neurodegenerative diseases [40].

Another data mining study implemented a new computational framework, including the use of the DBSCAN algorithm and Limma statistical methods, to analyze GEO datasets and identified 21 and 89 differentially methylated genes for AD and Down syndrome respectively [41]. Their evaluation indicated high classification accuracy of these two methylation signatures with 92% for AD and 70%

for Down syndrome. Their framework is capable of detecting outlier-free epigenetic signatures in complex diseases, with applications to analyze various epigenetic signatures throughout disease pathogenesis [41]. Studies performing meta-analyses of epigenomic datasets have found differentially methylated genes in varying brain regions [41], as well as age-associated methylation patterns concurrent with epigenetic dysregulation observed in AD [42].

It is important to note that epigenetic features and RNA expression as well as the subsequent protein expression level do not necessarily have a direct correlation [43]. With most drugs targeting proteins, it is vital to take a systems bioinformatics approach and integrate multiple types of omics data to holistically understand the disease mechanisms and regulation of protein synthesis at all levels of the central dogma [44]. There is also a need to characterize epigenetic changes at the cell-specific level with spatial resolution [45–49]. Going forward, personalized medicine aimed at targeting specific epigenetic changes in patients with neurodegenerative diseases is poised to drastically transform diagnostic and therapeutic strategies [50].

## 2.2. Bulk and single-cell/single-nucleus RNA sequencing analysis

One of the most commonly used methods in transcriptomics for gaining insight into which genes are differentially expressed in varying disease states is the RNA sequencing (RNA-seq). Transcriptomics analysis has been used to produce data representative of mRNA expression levels of tens of thousands of genes, as well as the identification of the differentially expressed genes (DEGs) in various biological samples between patients and healthy controls. RNA-seq allows for the sequencing of the whole transcriptome [51] and provides additional information on splice variants or non-coding RNA [52,53]. The databases that archive RNA-seq datasets include GEO [38], Mount Sinai Brain Bank [54], ROSMAP database [55], ARCHS4 [56], DDBJ Sequence Read Archive (SRA) [57] and Synapse database [58] (Table 1, Bulk RNA-seq database).

A data mining study focusing on biomarker discovery for AD utilized RNA-seq datasets stored in the DDBJ SRA and identified the gene *NEUROD6* to be downregulated in the brain tissue of AD patients [59]. Another study utilizing RNA-seq datasets obtained from the ROSMAP database and Mayo Clinic studies found that disease pseudotime (an arbitrary unit of time to measure a cells progression) in AD is significantly concordant with the burden of tau, A $\beta$ , and cognitive diagnosis of late-onset AD [60]. Additionally, it was reported that early-stage disease pseudotime samples show changes in basic cellular functions, while the late stage disease pseudotime samples show changes in neuroinflammation and amyloid pathologic processes [60].

Another data mining study uses the DDBJ SRA to obtain RNA-seq datasets from 26 different studies involving brain tissues and blood samples of AD and PD patients for meta-analysis [61]. By applying a random forest-based machine learning algorithm to analyze existing central and peripheral transcriptomic data, it was found that there is little overlap between AD and PD. Interestingly, the study revealed an overlap between central and peripheral transcriptomic signatures in PD that are characterized by anomalies in exocytosis and specific genes related to the SNARE complex including vesicle-associated membrane protein 2 (*VAMP2*), syntaxin 1A (*STX1A*), and p21-activated kinase 1 (*PAK1*) [61]. In a separate PD study making use of RNA-seq datasets from the GEO database, several genes including *RPL21*, *RPL34*, *CKS2*, *B2M*, *TNFRSF10A*, *DTX2*, and *HLA-B*, have been shown to be upregulated in PD brain tissues [62]. In MS, a data mining study analyzing RNA-seq datasets from the GEO database has reported that inactive MS brain lesions contain significantly more M2 macrophages compared to normal white matter controls [63].

Besides bulk RNA-seq, studies using single-cell RNA-seq (scRNA-seq) and single-nucleus RNA-seq (snRNA-seq) techniques are of high interest because they provide not only the average expression level for an ensemble of cells such as in the typical RNA-seq analysis [64,65], but also the ability to quantify gene expression levels in specific cell types [64–68]. Both techniques provide greater depth and insight into the analyzed data when compared to bulk RNA-seq. The scRNA-seq and snRNA-seq approaches are important in revealing cell subpopulations and intercellular heterogeneity, understanding regulatory relationships between genes, and tracking trajectories of distinct cell lineages in development, to understand disease pathogenesis [64,65]. In particular, the ability of scRNA-seq and snRNA-seq techniques to dissect the functional changes of highly heterogeneous cells in the brain at the single-cell level can significantly improve our understanding of the vulnerability of particular cell types in certain neurodegenerative diseases [69]. Some of the common databases and repositories used for data mining of scRNA-seq/snRNA-seq datasets include ARCHS4 [56], DDBJ SRA [57], Synapse database [58], Allen Brain Map [70], DRscDB database [71], and scREAD database [69] (Table 1, sc/snRNA-seq database).

A study making use of scRNA-seq datasets from scREAD to analyze the entorhinal cortex of AD brains found that phosphoinositide 3-kinase (PI3K)/protein kinase B (AKT) signaling, Wnt signaling, neuroactive ligand-receptor interaction pathways, and neurodegeneration pathways were significantly impaired in astrocytes from the entorhinal cortex of AD patients [72]. A similar data mining study using scRNA-seq data in combination with bulk RNA-seq data obtained from the Synapse database found significant upregulation of *PLCG2* expression in AD patients which positively correlates with amyloid plaque density [73]. This finding was validated by using an AD mouse model which showed increased *PLCG2* expression associated with amyloid pathology and disease progression and reducing microglia reverses the disease pathology [73]. A different study analyzed three snRNA-seq datasets [74–76] and showed that *LINGO1* is upregulated in both excitatory neurons and oligodendrocytes, together with indication of mitochondrial and estrogen signalling dysfunction in AD [77].

Finally, a study integrated both scRNA-seq datasets from scREAD and bulk RNA-seq datasets obtained from the Mount Sinai Brain Bank and ROSMAP databases for a comprehensive drug repositioning analysis [78]. They identified multiple new candidates for AD treatment such as trichostatin, which was predicted to be broadly applicable to different AD subtypes, and vorinostat, which was specific for one subtype of AD, and both of which are histone deacetylase inhibitors [78]. It is important to note the lack of scRNA-seq and snRNA-seq data mining studies for PD and MS which might be due to the limited number of databases dedicated to compiling scRNA-seq/snRNA-seq data for these diseases. While providing greater depth, scRNA-seq and snRNA-seq are lacking the spatial information which is achievable with image-based transcriptomics which we will discuss in the subsequent section.

## 2.3. Proteomics analysis

Functional proteins and their interactions with other molecules are essential for biological processes and cellular functions which govern the disease mechanisms in neurodegenerative diseases [79]. Importantly, translation of proteins from RNA is not a linear relationship, where some genes may not even translate into functional proteins [80–82]. Therefore, it is important to characterize protein expression in addition to RNA quantification when investigating disease states [79]. Proteomics provides additional biological insights such as protein-abundance differences in proteomes, time-dependent expression patterns, post-translational



modifications, and protein-protein interactions (PPIs) that otherwise could not be obtained from transcriptomics [79]. These parameters have been shown to have paramount effects on the pathogenesis and progression of neurodegenerative diseases [83].

Proteins rarely act as isolated machinery, and rather, their functionality is highly related to the proteins they interact with. Therefore, proteins whose function is well understood may be used to predict the function of unidentified proteins [84]. Both experimental and computational studies of PPI has enabled and expedited the modelling of functional pathways to elucidate the pathogenic mechanisms of cellular processes and identify their translational applications [85,86]. There is currently a lack of data mining of proteomics studies, although there have been efforts to compile mass spectrometry proteomics datasets into databases such as Global Proteome Machine Database (GPMDB) [87], jPOSTrepo [88], MassIVE [89], Proteomic Data Commons [90], ProteomeXchange [91], PRoteomics IDentifications Database (PRIDE) [92], and ProteomicsDB [93] (Table 1, Proteomics database). ProteomicsDB is an example of a database that allows users to explore and retrieve protein abundance values across different tissues, cell lines, and body fluids [93].

A neurodegenerative disease focused data mining study compiled proteomics studies of post-mortem brain tissue and used a meta-analysis approach to discover that biological processes related to the organization of the extracellular matrix, metabolism of glycosaminoglycans and proteoglycans, blood coagulation, response to injury, and oxidative stress were highly dysregulated in AD, PD, and Huntington's disease through PPI network analysis and Gene Ontology (GO) enrichment analysis [94]. Another study utilized the PRIDE database to study post-translational modifications in AD patients. They report 103 proteins with post translational modifications that are uniquely expressed between brain region with no tangles, intermediate tangles, and severe tangles [95]. The bioinformatics analysis suggested the association of these proteins in AD progression through platelet activation, and they were found to be enriched for the tricarboxylic acid cycle (Kreb's Cycle), respiratory electron cycle, and detoxification of reactive oxygen species [95]. Another proteomics study making use of meta-analysis found that pathways related to synaptic signaling, oxidative phosphorylation, immune response, and extracellular matrix were commonly dysregulated in AD through bioinformatic gene set enrichment analysis with the Enrichr web server [96,97].

Proteomic expression data provides insight into the involvement of post-transcriptional editing and quantifies protein encoding mRNA genes that make it through translation, and ultimately play critical roles as functional proteins [80,81,98]. As bioinformatics analysis is becoming more refined and expansive, it is vital for technological advances to keep up with the growing need for specificity. The importance of using combinatorial methods of transcriptomics and proteomics is now well established [99,100]. Single-cell proteomics analysis techniques are now bridging the gap by filling the growing need of specificity in understanding cell heterogeneity in neurodegenerative disease tissues [99]. Single-cell proteomics makes use of mass spectrometry techniques for proteome quantification by analyzing individual cells one at a time [99,100]. This new and growing field will give novel insights into cell heterogeneity in tissue samples, as well as providing information into transcriptional regulation of DEGs when compared to scRNA-seq datasets [82,99,101].

#### 2.4. Metabolomics/lipidomics analysis

Metabolomics and lipidomics were once classified under the same umbrella, but each of them now occupies an independent

domain due to the large range of studies characterized by each of the methods extensively [102]. Metabolomics focuses mostly on examining polar metabolites such as sugars, amino acids, organic acids, and nucleotides which are usually the end products of complex biochemical cascades. On the other hand, lipidomics strives to identify lipid molecular species which should be analyzed separately from small-molecule metabolites due to their hydrophobic nature [103,104]. Importantly, with the brain being the second most abundant organ in terms of lipid concentration and diversity, lipid dysregulation has been largely linked to AD, PD, and MS due to the vital tasks of lipids in myelination of neurons and signal transduction via lipid mediators [105].

Metabolomics and lipidomics datasets are typically acquired from similar biological samples using common analysis methods such as mass spectrometry, ion chromatography, liquid chromatography, and nuclear magnetic resonance [103,106,107]. Data collected from metabolomics and proteomics are not mutually exclusive to each other, due to the intertwined relationship they both have to biological processes involved in cellular homeostasis and pathogenesis of neurodegenerative diseases, which will contribute to potential diagnosis and therapeutic targeting of these diseases [108,109]. With the expanding experimental data being collected, there are now many openly accessible metabolomics and lipidomics databases such as Alzheimer's Disease Neuroimaging Initiative (ADNI) database [110], Cerebrospinal Fluid (CSF) Metabolome Database [111], Human Metabolome Database (HMDB) [112–116], Lipid Bank [117], LipidBlast [118], Lipid MAPS [119], MetaboAge Database [120], MetaboLights [121], Metabolomics Workbench [122], MetabolomeXchange [123], and Serum Metabolome Database [124] (Table 1, Metabolomics/Lipidomics database).

In terms of lipidomics, a data mining study utilized datasets obtained from the ADNI database [110], which consisted of 349 serum samples obtained from 806 participants, to investigate lipid metabolism in AD [125]. They found lipid desaturation, elongation, and acyl chain remodeling processes to be disturbed in the blood of AD patients. The study further tested the association between sets of blood lipids with known AD biomarkers and showed that A $\beta$  in the CSF correlates with glucosylceramides, lysophosphatidylcholines, and unsaturated triacylglycerides [125]. On the other hand, there is a scarcity of metabolomics-based data mining studies associated with neurodegenerative diseases. To investigate aging and associated diseases, MetaboAge database has compiled metabolomics data from dozens of studies reporting statistically significant changes in metabolites associated with ageing in healthy individuals [120]. This database may serve as an informative platform to compare metabolic changes between ageing and the mechanisms of neurodegenerative diseases obtained from other databases to facilitate future data mining studies using metabolomics datasets.

As examples, there are several data mining studies utilizing metabolomics datasets to understand the changes in metabolites in other neurological diseases such as glaucoma and depression. A meta-analysis study looking at primary open angle glaucoma (POAG) identified aminoacyl-tRNA biosynthesis and arginine metabolisms, which play important roles in immune responses, being dysregulated in patients with POAG compared to controls [126]. Another study examining altered metabolites in depression compiled 5,675 metabolite entries from 464 studies collected from metabolomic databases, including HMDB, MetaboLights [121], Metabolomics Workbench [122], MetabolomeXchange [123], as well as Omics Discovery Index [127] which contain all omics datasets, together with extensive literature survey [128]. They found that patients with depression had lower levels of brain gamma-aminobutyric acid and glutamate/glutamine, and that

tryptophan metabolism-related metabolites such as serotonin, 5-hydroxyindoleacetic acid, quinolinic acid, and tryptophan were most frequently changed after treatment [128].

### 2.5. Spatial omics analysis

Spatial omics technologies have provided new opportunities to visualize the anatomical localization of biological molecules to enable the investigation of the structural organization of complex tissue as well as visualization of the interactions between cells and their tissue microenvironments [129–132]. While spatial analysis has been increasingly applied to all types of omics studies, most of the current studies focus on spatial transcriptomics and proteomics analyses. A number of technological advances have enabled transcriptomics and proteomics profiling where the transcripts or proteins can be assigned to their specific cell types and cell location [133–136], revealing distinct spatial patterns of cells in tissues that were previously inferred through indirect means [137–139]. Techniques used for spatial transcriptomics analysis include fluorescence in situ hybridization (FISH) [140], seqFISH+ [136], and multiplexed error-robust FISH (MERFISH) [133], while spatial proteomics analysis makes use of techniques such as cytometry by time of flight (CyTOF) [141] and highly multiplexed immunofluorescence imaging approaches [142]. All analytical techniques have been applied to analyze AD [76,143–145], PD [146–148], and MS tissues [149–153].

To facilitate the interpretation of spatial transcriptomics and proteomics datasets, interactive visualization tools are typically used, including SpatialLIBD [154], SpatialExperiment [155], Bento [156], MSnbase [157], pRoloc [157], Squidpy [158], Spatial Multi-Omics (SM-Omics) [159], ATHENA [160], and TRANSPIRE [161,162]. These analysis tools process single-cell transcriptomics and proteomics data and computes spatial statistics of subcellular RNA and protein molecular distributions, compartmental expression, and cell morphology to build multidimensional biological features associated with diseased states as compared to controls [156]. Databases and repositories can further facilitate the comprehensive archiving and exploration of spatial omics datasets to enable data mining or comparisons with other experimental data [163,164]. Commonly used databases include PRIDE [92], ProteomicsDB [93], Dynamic Proteomics [165], Giotto [166], Spatial TranscriptOmics DataBase (STOmicsDB) [167], and SpatialDB [168] (Table 1, Spatial Omics database). For example, SpatialDB contains functions such as the ability to search for relevant publications and tools, public dataset visualization, customized specialized databases, new data archive, and online analysis [168]. It is important to note that while there are several analytical tools and databases available to aid in spatial transcriptomic and proteomics data analysis, not many studies have made use of such tools for data mining.

A prominent example is Giotto, which utilizes a rich variety of algorithms that enables robust spatial data analysis, and a user-friendly platform for data visualization and exploration, including characterizations of tissue composition, spatial expression patterns, and cellular interactions [166]. Giotto has been shown to be applicable to a wide range of public datasets, including several spatial datasets from neurodevelopment and neurodegeneration studies that illustrated consistent analysis and conclusion [166]. In addition, RNA-seq data can be integrated for spatial cell-type enrichment analysis [166,169,170]. Giotto, for example, has utilized single-cell spatial transcriptomic MERFISH data that was collected from the pre-optic cortex of a mouse, and was able to identify 8 distinct cell clusters as well as creating interactive three-dimensional plots of the dataset [166] where the results are concordant with the original study from which the data was obtained. Giotto has additionally been used to predict the presence of

a given cell type in a spatial location with multiple cell types for datasets with low spatial resolution. The spatial cell type prediction algorithm was tested by altering a seqFISH + dataset to mimic low spatial resolution. The cell-type enrichment analysis was conducted by using scRNA-seq data and derived marker gene lists for somatosensory cortex associated cell types obtained from a previous study [171], where there is a high accuracy in predicting the presence of a cell type at each individual spatial location [166].

A data mining study made use of a multi-modal structured embedding (MUSE) approach to analyze five datasets consisting of seqFISH+, STARmap, spatial transcriptomics, Visium, and spatial transcriptomics with fluorescent imaging data [172]. Application of MUSE to these diverse datasets yielded spatial patterning in healthy mice brain cortex tissue utilizing seqFISH + data, as well as heterogeneity of amyloid precursor protein processing in mice AD brain regions [172]. MUSE also successfully clustered STARmap mouse cortex data and differential expression analysis allowed for the identification of the clusters as astrocytes, hippocampal neurons, oligodendrocytes, or smooth muscle cells. Using transcriptomic and immunofluorescent imaging of AD mice brain tissue, MUSE was able to identify DEGs in individual clusters [172]. It was found that known AD-related genes *RANBP9* (downregulated in hypothalamus), *IGF1* (upregulated in cortex), and *SORL1* (upregulated in hypothalamus; downregulated in cortex) were differentially expressed [172]. This approach has revealed regional, temporal, and biological differences reflecting AD progression in a mouse model [172].

Spatial transcriptomics and proteomics have substantially advanced our ability to detect the heterogeneity of RNA and protein expression in tissues, although characterizing whole-transcriptome data of individual cells in space remains a challenge [173]. Integrating spatial transcriptomics with scRNA-seq and snRNA-seq techniques has been on the rise with hopes of resolving the limitations that spatial transcriptomics currently pose by gaining a deeper understanding of cell-cell communication within healthy and disease tissues and the roles certain cell subpopulations have in maintaining homeostasis and disease pathogenesis [174–177]. With the potential to incorporate other types of spatial omics datasets as they become available to advance towards spatial multi-omics [178], we will be a step closer to elucidate the detailed tissue organization, cell regulation, and cellular communication at an unprecedented scale.

### 3. Towards integrative multi-omics and systems bioinformatics to reconstruct brain profiles

Neuroscience is being propelled into the big data era with an exponential increase in the amount of information generated which demands for better data organization, improved pipeline frameworks, and rapid turnover for analysis and interpretation to turn this information into valuable biological insights [179]. A systems bioinformatics based data mining approach enables the integration of a spectrum of multi-omics information ranging from epigenetics/epigenomics, transcriptomics, and proteomics to metabolomics and lipidomics by a combination of data-driven bioinformatics (top-down approach) and systems biology (bottom-up approach) [14]. It has also been proposed that establishing multiple networks representing information obtained from each type of omics dataset and integrating them in a layered network that exchanges information within and between layers could enable the comprehensive systems bioinformatics analysis (Fig. 1) [14]. One of the main challenges of integrative approaches is related to increased dimensionality, due to increased complexity of the omics dataset in the biological system. Experimentally, several neurodegenerative focused studies have started to incorporate a multi-omics approach in their analyses [180–184].

In both individual and multi-omics analyses, after data mining of the omics datasets, the data must be transformed and processed through normalization, quality control, and feature selection to extract interpretable information (Fig. 2). Normalization is typically applied to most omics layers to remove bias, large variation, and outlier or incorrect reads in order to make better comparison between different datasets or omics layers [185]. Quality control should be applied to all omics layers such as through quantifying GC content in RNA-seq and removing duplicated and fragmented reads for sequence alignment to be performed [186]. Feature selection is commonly conducted to reduce the dimensionality and redundancy of the high-throughput data, and discriminate desired features contained within the data [187,188]. The goal of the data processing is to reduce dimensionality, bias, and variation of the mined data in order to ensure robustness and efficiency of analysis, especially prior to multi-omics integration. Next, we describe two multi-omics integration approaches in data mining, namely independent biological integration and unsupervised integration, to combine individual layers of omics data for an integrative multi-omics analysis (Fig. 3A).

### 3.1. Independent biological integration

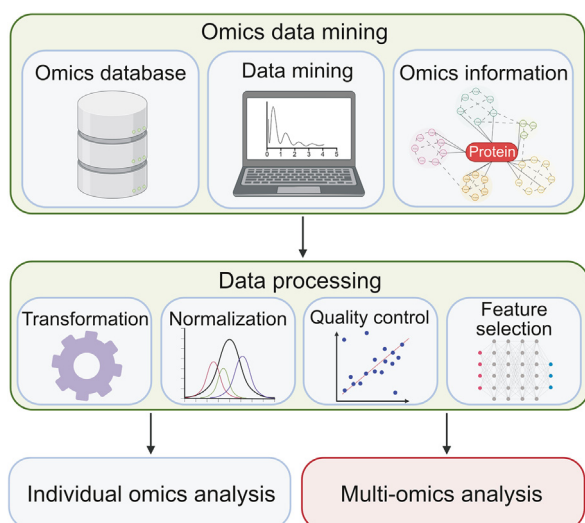
In independent biological integration, different types of omics datasets are typically analyzed by isolating the DEGs at the individual layers first and then compiled for integrated analysis. Integration is then performed by biological intuition such as comparing expression levels of genes and proteins to understand the translational regulation [189]. Integration can also be done through computational based or web-based tools that integrate expression levels of various types of omics data for biological interpretation, functional annotations, gene set enrichment, and multiscale network-based visualization to understand the interplay of regulation at different levels of gene expression [190,191] (Fig. 3B). Some of the computational tools include Kyoto Encyclopedia of Genes and Genomes (KEGG) [192,193], GO [194,195], DAVID [196], PANTHER [197,198], GSEA [199], and IPA [200] for pathway and gene enrichment analyses. The PPIs in functional modules and their

interactions with each other in cellular networks can be determined by STRING [201], Cytoscape [202], NPA [203], and SPIA [204]. It is important to note that there might be variability between these computational tools, and it is advisable to use these computational tools in combination to define the most relevant and highly reproducible pathways and networks associated with certain disease states. To ensure robustness and accuracy of results, there must be consistency in parameters used for statistical analysis in each individual omics layer.

Other web-based tools such as MetExplore [205], 3Omics [206], and PaintOmics [207] have started to layout the potential of multi-omics analysis and visualization based on the independent biological integration approach [208]. The MetExplore program enables the visualization and interpretation of omics data from multiple molecular layers with inputs from different omics data followed by providing an interpretation of genome-scale metabolic networks and how various types of omics data modulate metabolic processes [205]. While the 3Omics program builds correlation networks, and enables phenotyping based on data from different omics layers [206], the PaintOmics programs allow for network visualization and accepts a wide variety of omics data [207]. The PaintOmics program additionally allows for pathway and enrichment analyses based on KEGG, Reactome, and MapMan databases. Using independent biological integration, omics datasets can be analyzed from multiple samples and do not need to be isolated from the same experiment. More complex forms of independent biological integration include “horizontal” (with features as anchors), “vertical” (with cellular information as anchors), and “diagonal” (no anchor) integration that are used for data mining of multi-modal single-cell datasets with anchors aligning different types of omics data [209]. Independent biological integration is currently being used as the primary method of omics integration in data mining studies due to ease of implementation without the need for highly technical computational competencies.

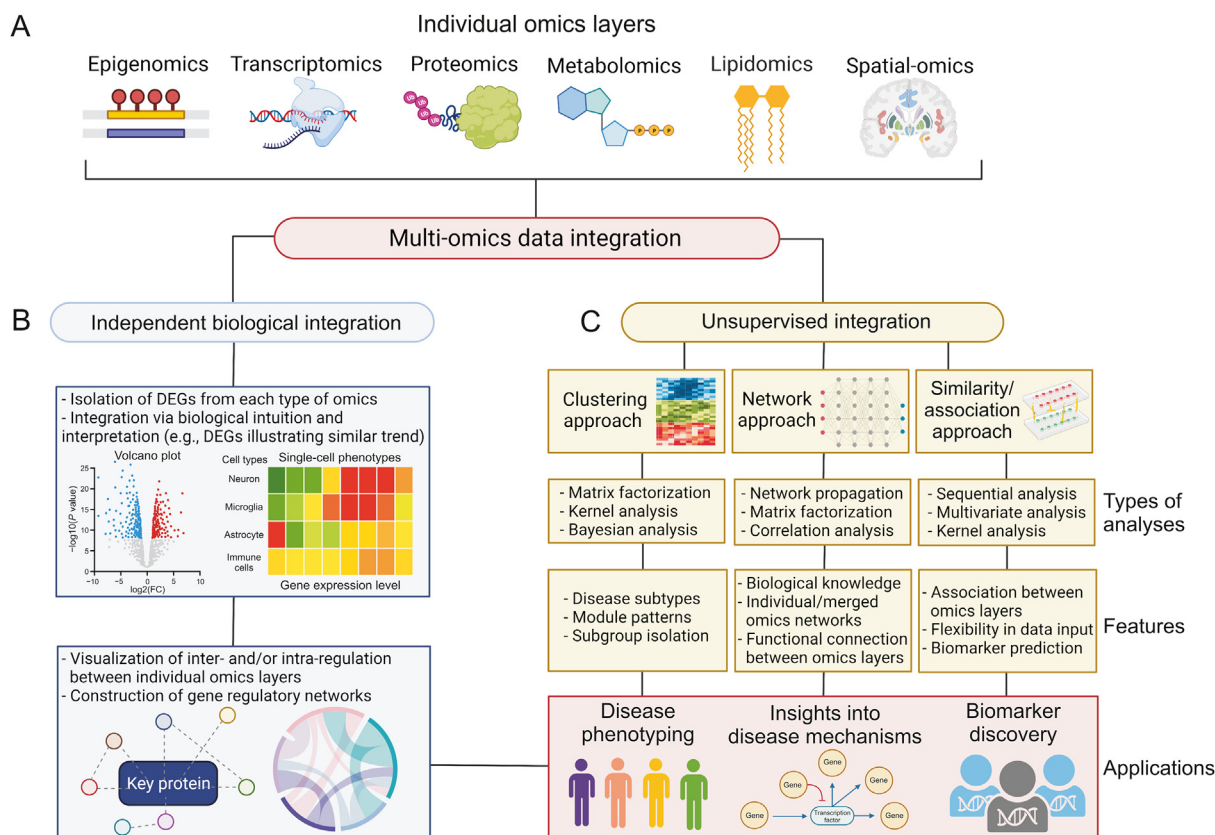
With the emergence of databases making multiple types of omics data publicly accessible, proof-of-concept integration of multi-omics datasets has been illustrated in data mining using the independent biological integration approach. A data mining study utilizing transcriptomic and proteomic datasets found significant DEGs in the human spinal cord of MS patients and the implications of these DEGs on biological processes involved in the disease progression of MS [189]. Specifically, they found that *HOXA5* was significantly upregulated in MS patients through individual transcriptomic (ARCHS4 database [56]) and proteomic analysis (BioGrid [210]) and that *HOXA5* was found to promote the transforming growth factor (TGF)-beta pathway [189]. A previous study has shown that in the spinal cord of MS patients there are large areas of demyelination characterized by a unique *TGFB1* genomic signature [189]. This study proposes that the overexpression of *HOXA5* in the spinal cord may promote the progression of *TGFB1*-mediated gliosis in MS patients [189].

Multi-omics integration has also been reported to be conducted with the help of the MetExplore program to process genome-wide association studies (GWAS), transcriptomics, and proteomics datasets obtained from several databases (GWAS catalog [34], GEO [38], and PRIDE [92], respectively) to extract differentially expressed multi-omics elements [190]. This study identified 203 differentially expressed transcripts, 164 differentially expressed proteins, and 58 differentially expressed GWAS-derived mouse orthologs associated with significantly enriched metabolic biological processes [190]. Additionally, lipid metabolic pathways were significantly upregulated across the multi-omics datasets, with microglia and astrocytes expressing significant enrichment in the lipid-predominant AD-metabolic transcriptome [190]. This study brings attention to the significance of dysregulated lipid metabolism in AD, and the



**Fig. 2.** Omics data mining and data processing for individual and multi-omics analysis. Omics data mining involves the exploration of various omics databases to search for suitable datasets to be used, mining of the data, and obtaining the pre-processed omics information. The data must then be transformed and processed through normalization, quality control, and feature selection to extract interpretable information. The figure was created with BioRender.com.





**Fig. 3.** Multi-omics data integration of individual omics layers. (A) Multi-omics data integration is typically performed through two main approaches, namely the independent biological integration and the unsupervised integration. (B) The independent biological integration approach involves the isolation of differentially expressed genes (DEGs) from each omics layer followed by integration via biological intuition and interpretation as well as network visualization and construction. (C) The unsupervised integration involves three main categories of analysis which includes clustering, network, and similarity/association approaches. Each approach makes use of various statistical, network, correlation, and sequential analyses as well as incorporates different features for the integration. Both independent biological integration and the unsupervised integration can provide insights into disease phenotyping, disease mechanisms, and biomarker discovery. The figure was created with [BioRender.com](https://BioRender.com).

importance and usefulness of using multi-omics analysis to better understand AD pathogenesis from a systems bioinformatics approach, with experimental metabolomics/lipidomics validation in a mouse model and in the blood plasma of AD patients, respectively [190]. Although there are several advantages using the independent biological integration approach, it is worth noting that different studies that make use of this method may be subject to different biological intuitions in the integration process, leading to non-standardized analysis and less consistent results.

### 3.2. Unsupervised integration

For unsupervised integration, multi-omics data from the different molecular layers are often concatenated together into a single matrix for analysis via ensemble dimension reduction [211,212]. Other approaches, such as model-ensemble, each omics layer is analyzed independently to obtain the respective matrix and the matrices from all omics layers are then inputted into the unsupervised algorithm and fused to build an integrated analysis. There are three main categories of unsupervised integration of multi-omics data, namely clustering-based, network-based, and similarity/association-based approaches (Fig. 3C). First, the clustering-based approach is primarily based on statistical calculations, making use of matrix factorization, kernel, and Bayesian analyses. Using matrix factorization analysis, the non-negative matrix factorization (NMF) method is most commonly utilized for high-dimensionality datasets and restricts their entries to non-negative values, allowing for easier interpretation of results [213].

Extensions of NMF include integrative NMF and joint NMF which account for the identification of heterogeneity and homogeneity in datasets respectively during the integration process [214,215]. Kernel analysis captures the degrees of similarity of the input data which are contained within the kernel matrix. This analysis is dimension-free and does not depend on the total number of features in the datasets. In the Bayesian analysis or Bayesian consensus clustering, a probability model such as the Dirichlet process mixture model, is used to model source-specific features as well as an overall clustering accounting for multiple data sources in different omics layers [216]. Clustering-based methods are suitable for identifying disease subtypes and module patterns, as well as isolating subgroups, samples, or features that have similar biological function.

The network-based unsupervised integration relies on biological knowledge databases for information on functional relationships between omics layers in addition to statistical analysis and calculation, and are heavily used to identify functional relationships between omics layers [212]. Network propagation analysis tracks the flow of each node and amplifies the signals based on the assumption that genes underlying similar phenotypes interact with one another through known information such as the association with a biological process [217]. Individual networks are then fused together into a similarity network using a nonlinear fusion approach which is based on message-passage theory [218]. Similar to the clustering-based method of analysis, network-based methods also utilize matrix factorization statistical approaches. On the other hand, correlation analysis is based on the correlation between a node, such as a gene and an outcome, with the



significance of a node determined by the correlation coefficient or a regression-based significance [219].

Finally, the similarity/association-based unsupervised integration approach enables the identification of the marginal associations and correlations between various omics layers. Sequential analysis is an example, where statistical tests and models are applied to narrow down the list of features in one omics layer based on their relationship with features in other omics layers [212]. Multivariate analyses including canonical correlation analysis (CCA) and co-inertia analysis (CIA) are useful methods due to its flexibility in accepting multiple matrices as input data. While CCA can be applied for feature selection and classification in high-dimensional multi-omics datasets, CIA is used to find the low-dimensional components and aims to distinguish sources of variation in multi-omics datasets [220,221]. Similarity/association-based methods can also utilize kernel statistical analysis similar to clustering-based methods. Similarity/association-based unsupervised integration approach enables biomarker prediction, associations between omics layers (e.g., genotypes based on gene expression) and flexibility in accepted data (e.g., multiple matrices).

There are also some commonly used programs that can be utilized to facilitate unsupervised integration such as iCluster programs [222,223], JIVE [224], CNAMet [225], and PARADIGM [226]. Briefly, the iCluster programs and JIVE are all matrix-factorization-based clustering methods used for disease subtyping. The iCluster programs create flexible models based on the associations between omics layers and determines the variance-covariance within omics layers in a single framework, all while simultaneously reducing dimensionality of each omics layer [222]. The JIVE program is an extension of principal component analysis (PCA) and calculates the amount of joint variation between omics layers, reduces dimensionality, and enables visual exploration of joint and individual structures such as patterns of biological relationships between omics layers [224]. The CNAMet program is a similarity/association-based sequential analysis tool that integrates high-throughput copy number, DNA methylation, and gene expression data that is used for biomarker prediction [225]. The PARADIGM program is a Bayesian-based network integration tool used for disease mechanistic studies and subtyping. It integrates multi-omics data to infer the modulation of genetic pathways based on established knowledge of the given pathways [226].

Unsupervised integration of multi-omics datasets has been mainly used in primary research, including AD, PD and ALS studies [4,12,227,228]. Recently, this approach has been adopted by neurodegenerative disease related data mining studies. For example, a study utilizes a combination of proteomics and lipidomics datasets collected from the blood of 586 AD patients and controls from other studies [229,230] for multi-omics analysis using the unsupervised integration approach [231]. Network analysis of the individual omics datasets was conducted using the Weighted Gene Correlation Network Analysis (WGCNA) which can also be applied to proteomics and lipidomics data [232,233]. Data processing which includes normalization, imputation of missing values, and PCA were performed before creating the weighted co-expression networks via hierarchically clustering and module assignments using a dynamic tree-cutting algorithm. To integrate the protein and lipid modules and analyze the associations between AD-associated modules, Pearson's correlation coefficient was utilized [231]. GO enrichment analysis was then used to analyze the biological processes as well as molecular and cellular functions of the protein modules associated with AD phenotypes. The study identified lipid modules involved in immune response and lipid metabolism as well as protein modules involved in increased cytokine production, humoral immune responses, and neutrophil-mediated immunity, all of which were highly correlated with AD risk loci [231]. This study is a good example

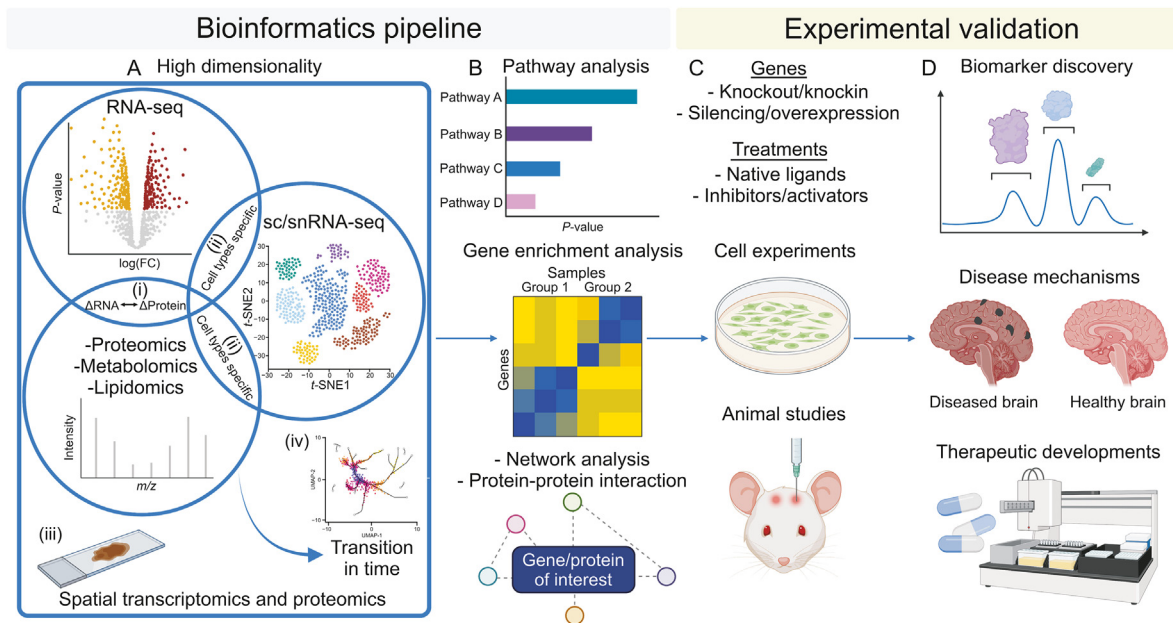
of an unsupervised multi-omics integration approach via data mining which exemplifies network-based approaches for isolating protein and lipid modules that are highly associated with established AD risk loci. Another example related to brain disease is a glioblastoma study making use of multi-omics datasets obtained from patients to develop a network-correlation-based method called Lemon-Tree for biomarker discovery [234]. This study demonstrated that the Lemon-Tree algorithm successfully identifies known oncogenes and tumor suppressors as master regulators in the inferred module network, utilizing somatic copy number and expression data. Lemon tree allows the addition of other omics features such as miRNA and DNA methylation to be added to the model, and for GO enrichment analysis of the modules.

Currently, the largest challenges with overcoming interoperability of omics data is the lack of a standardized framework or pipeline, to enable various types of omics data to be seamlessly integrated and analyzed [208,212]. Systems bioinformatics extracts disease relevant information from multiple levels of the omics spectrum and integrates them in a layered and interactive network that exchanges systemic knowledge towards developing a comprehensive brain profile [235–237]. Although the process of optimizing the derived results is certainly required to obtain reliable biological information, this approach is theoretically applicable to the whole omics spectrum to work towards achieving systems bioinformatics in translational neuroscience. Our proposed concept focuses on a broad idea of obtaining ultimate research goals of reconstructing human brain systems with direct healthcare relevance [238], rather than a detailed in-depth analysis pipeline or algorithm development to analyze specific datasets [239–241]. We note here that highly technical methodologies, mathematical algorithms, and information theory are necessary to understand the omics derived networks.

#### 4. Data mining and experimental validation in translational neuroscience

Data mining approaches have advantages of being high throughput and low cost as compared to traditional low throughput experimental techniques in revealing the underlying pathogenic mechanisms of complex neurodegenerative diseases. However, prediction results arising from data mining remain theoretical and require validation with experimental evidence [242]. Here, we recommend an integrative neuroscience approach that synergizes systems bioinformatics and experimental analysis to yield opportunities for translational applications including biomarker discovery, therapeutic development, and insights into disease mechanisms (Fig. 4).

In the previous section, we have described multi-omics approaches to integrate different individual omics data together for a more holistic interpretation of the results. Besides omics data, high dimensionality analysis should include cell type, spatial location, and time trajectory (Fig. 4A). The knowledge of the biological functions and network systems in the brain such as the key pathways, genes, and PPI involved are important to piece all information together to provide interpretable biological findings (Fig. 4B). Under the context of multi-omics and systems bioinformatics analysis, one would expect consistent or correlated alterations in different omics layers, be it localized information or circulating information, to enable the reconstruction of the brain profile and understanding of the brain physiology. Nuanced analysis might be encountered which requires further optimization in data analysis and confirmation through experimental validation. Experimental validation bridges the gap between bioinformatics analysis and translational neuroscience and increases the credibility of results, especially for novel discoveries. For example, it is important to test how the



**Fig. 4.** Proposed bioinformatics pipeline with experimental validation for translational neuroscience applications. (A) The high dimensional omics analysis includes (i) the overlap between transcriptomics and proteomics/metabolomics/lipidomics datasets which identifies the differentially expressed genes (DEGs) that are altered across all omics layers; (ii) the overlap with single-cell/single-nucleus RNA-sequencing (sc/snRNA-seq) indicates the identification of DEGs in different cell types or cell subpopulations; (iii) the overlap with spatial analysis enables the understanding of the location of the DEGs in the respective brain regions; and (iv) the overlap with data with time element will provide further information on how the DEGs and the corresponding phenotypes could change over time. (B) Key cluster of DEGs can be further analyzed by pathway analysis, gene-enrichment analysis, network analysis and/or protein-protein interaction analysis, among many other analyses that can be performed. (C) The key DEGs should be validated by cell experiments or animal studies with strategies such as gene knockout or knockin, gene silencing or overexpression, or treatments with native ligands or protein inhibitors/activators. (D) Changes in gene/protein levels or alteration in protein activities associated with certain disease phenotypes may provide insights to disease mechanisms, as well as the identification of potential biomarkers or advancement of potential therapeutic developments. The figure was created with [BioRender.com](https://www.biorender.com).

overall biological system of cells and/or mice would react to alterations in certain key protein achieved by either overexpression/knockout or treatment of protein ligands/modulators (Fig. 4C). It is also important to examine whether these alterations have any adverse consequences on the model organisms tested such as toxicity. Generally, a protein plays a key role in disease mechanism if the change in the protein level or function correlates with disease pathogenesis or progression. Therapeutic discovery can be achieved through screening of small molecules or antisense oligonucleotide that can modulate the protein function. A biomarker is typically established if alterations in certain key proteins that are associated with disease pathogenesis can be detected in blood or CSF prior to disease progression (Fig. 4D).

Overall, it is essential for computational scientists, experimental biologists, and neuroscientists to communicate and exchange the intricacies of their individual methods to enable evaluation and validation of results with appropriate interpretation in a balanced manner [243]. Integrative neuroscience, combining both systems bioinformatics and experimental biology analysis, is a multidisciplinary science that can provide a new approach for biomarker discovery, therapeutic development, and elucidation of pathogenic mechanisms of neurodegenerative diseases [239].

### 5. Summary and future perspectives

Bioinformatics is becoming increasingly essential for the organization and management of data in modern biology and is a comprehensive field that harnesses methods in computer science and experimental biology. Bioinformatics is redefining modern science in ways that were not possible in the past with the ability to combine datasets from multiple experiments and different samples for large analysis has never been as accessible and accurate as today, yielding ever-growing applications of bioinformatic analyses

[244]. Systems bioinformatics is a rising concept which makes use of network-based computational analysis to increase the precision of mechanistic understanding of disease pathogenesis as well as development of new diagnostics and therapeutics [14].

This review accentuates the importance of understanding the applications and boundaries of various data mining approaches of omics datasets and computational methods towards multi-omics analysis. The integration of multi-omics analysis, systems bioinformatics, and experimental validation provides insights into disease mechanisms and opens avenues for translational neuroscience applications as well as advancing towards early diagnosis [245], precision phenotyping of diseases [246] and personalized medicine [247]. With a resolution of exactly understanding the molecular level changes of a certain proteins in a specific cell-type as well as its detailed location and trajectory in time, it is theoretically possible to target the exact cells with the pathological features and provide treatments, although this will require highly specific targeting strategies to be developed.

A number of limitations remain for systems bioinformatics and experimental analysis in translational neuroscience before it can be fully implementable [248,249]. A major challenge lies in validating the reconstructed molecular networks with real biological observations as there is a lack of ground truth and technical capabilities to recapitulate the layered networks in the brain [14]. It has been suggested that an initial step would be to compare reconstructed networks with benchmarking datasets with known biological information such as from existing curated databases and simulated datasets that mimic real data [250]. In addition, similar to the current limitation in data mining where there is no streamlined pipeline for bioinformatics analysis, there is also variability in the network reconstruction approaches, leading to inadequate network selection and inconsistent results between different studies [251]. Hence, there is a need for computational approaches to screen

results from different methods of analyses to provide a consensus analysis that will maximize the information content, although this remains to be developed.

Heterogeneity of neurodegenerative diseases within individuals is caused by multiple complex factors and is impeding development of effective treatments, leading to the notion of personalized medicine. Systems bioinformatics methodologies enable the collection of invaluable knowledge gathered from the different aspects of a disease condition and provide revolutionary approaches and tools to clinicians to demystify the complex nature of these diseases. Although the full implementation of the systems bioinformatics approach to reconstruct the human brain profile might seem ambitious at the current stage, its current application might be complementary with the use of existing computational and translational neuroscience methods. The practical integration of systems bioinformatics and experimental analysis in translational neuroscience is likely to have a major impact and significant breakthroughs in detection and diagnostics of neurological disorders and neuroscience targeted drug discovery in the future [252], and may be applicable to studying other diseases in general.

### CRediT author statement

**Lance M. O'Connor:** Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - Original draft preparation; **Blake A. O'Connor:** Data curation, Writing - Reviewing and Editing; **Su Bin Lim:** Writing - Reviewing and Editing; **Jialiu Zeng:** Funding acquisition, Validation, Writing - Reviewing and Editing; **Chih Hung Lo:** Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Validation, Visualization, Writing - Original draft preparation.

### Declaration of competing interest

The authors declare that there are no conflicts of interest.

### Acknowledgments

Chih Hung Lo is supported by a Lee Kong Chian School of Medicine Dean's Postdoctoral Fellowship (021207-00001) from Nanyang Technological University (NTU) Singapore and a Mistletoe Research Fellowship (022522-00001) from the Momental Foundation USA. Jialiu Zeng is supported by a Presidential Postdoctoral Fellowship (021229-00001) from NTU Singapore and an Open Fund Young Investigator Research Grant (OF-YIRG) (MOH-001147) from the National Medical Research Council (NMRC) Singapore. Su Bin Lim is supported by the National Research Foundation (NRF) of Korea (Grant Nos.: 2020R1A6A1A03043539, 2020M3A9D8037604, and 2022R1C1C1004756) and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant No.: HR22C1734). The authors thank Jonathan Indajang from Cornell University for proofreading the manuscript.

### References

- [1] J. Gauthier, A.T. Vincent, S.J. Charette, et al., A brief history of bioinformatics, *Brief. Bioinform.* 20 (2019) 1981–1996.
- [2] V. Svensson, R. Vento-Tormo, S.A. Teichmann, Exponential scaling of single-cell RNA-seq in the past decade, *Nat. Protoc.* 13 (2018) 599–604.
- [3] D.H. Geschwind, G. Konopka, Neuroscience in the era of functional genomics and systems biology, *Nature* 461 (2009) 908–915.
- [4] X. Dong, C. Liu, M. Dozmorov, Review of multi-omics data resources and integrative analysis for human brain disorders, *Brief. Funct. Genomics* 20 (2021) 223–234.
- [5] Y. Hasin, M. Seldin, A. Lusic, Multi-omics approaches to disease, *Genome Biol* 18 (2017), 83.
- [6] B. Tasic, Single cell transcriptomics in neuroscience: Cell classification and beyond, *Curr. Opin. Neurobiol.* 50 (2018) 242–249.
- [7] E. Lein, L.E. Borm, S. Linnarsson, The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing, *Science* 358 (2017) 64–69.
- [8] W.X. Wang, J.L. Lefebvre, Morphological pseudotime ordering and fate mapping reveal diversification of cerebellar inhibitory interneurons, *Nat. Commun.* 13 (2022), 3433.
- [9] S.L. Wilson, G.P. Way, W. Bittremieux, et al., Sharing biological data: Why, when, and how, *FEBS Lett* 595 (2021) 847–863.
- [10] G.M. Shepherd, J.S. Mirsky, M.D. Healy, et al., The Human Brain Project: Neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data, *Trends Neurosci.* 21 (1998) 460–468.
- [11] C. Villa, J.H. Yoon, Multi-omics for the understanding of brain diseases, *Life* 11 (2021), 1202.
- [12] C. Clark, M. Rabl, L. Dayon, et al., The promise of multi-omics approaches to discover biological alterations with clinical relevance in Alzheimer's disease, *Front. Aging Neurosci.* 14 (2022), 1065904.
- [13] F.V. Greco, A. Pandi, T.J. Erb, et al., Harnessing the central dogma for stringent multi-level control of gene expression, *Nat. Commun.* 12 (2021), 1738.
- [14] A. Oulas, G. Minadakis, M. Zachariou, et al., Systems Bioinformatics: Increasing precision of computational diagnostics and therapeutics through network-based approaches, *Brief. Bioinform.* 20 (2019) 806–824.
- [15] S. Grillner, A. Kozlov, J.H. Kotaleski, Integrative neuroscience: Linking levels of analyses, *Curr. Opin. Neurobiol.* 15 (2005) 614–621.
- [16] T. Schneider-Poetsch, M. Yoshida, Along the central dogma-controlling gene expression with small molecules, *Annu. Rev. Biochem.* 87 (2018) 391–420.
- [17] G. Calabrese, C. Molzahn, T. Mayor, Protein interaction networks in neurodegenerative diseases: From physiological function to aggregation, *J. Biol. Chem.* 298 (2022), 102062.
- [18] C.H. Lo, J.N. Sachs, The role of wild-type tau in Alzheimer's disease and related tauopathies, *J. Life Sci. (Westlake Village)* 2 (2020) 1–17.
- [19] C.H. Lo, Heterogeneous tau oligomers as molecular targets for Alzheimer's disease and related tauopathies, *Biophysica* 2 (2022) 440–451.
- [20] C.H. Lo, Recent advances in cellular biosensor technology to investigate tau oligomerization, *Bioeng. Transl. Med.* 6 (2021), e10231.
- [21] Y. Hou, X. Dan, M. Babbar, et al., Ageing as a risk factor for neurodegenerative disease, *Nat. Rev. Neurol.* 15 (2019) 565–581.
- [22] L. Gan, M.R. Cookson, L. Petrucelli, et al., Converging pathways in neurodegeneration, from genetics to mechanisms, *Nat. Neurosci.* 21 (2018) 1300–1309.
- [23] R. Ghosh, S.J. Tabrizi, Gene suppression approaches to neurodegeneration, *Alzheimer's Res. Ther.* 9 (2017), 82.
- [24] I.A. Qureshi, M.F. Mehler, Advances in epigenetics and epigenomics for neurodegenerative diseases, *Curr. Neurol. Neurosci. Rep.* 11 (2011) 464–473.
- [25] G. Yu, Q. Su, Y. Chen, et al., Epigenetics in neurodegenerative disorders induced by pesticides, *Genes Environ.* 43 (2021), 55.
- [26] P. Ghosh, A. Saadat, Neurodegeneration and epigenetics: A review, *Neurologia (Engl Ed)* 38 (2023) e62–e68.
- [27] F. Coppede, Targeting the epigenome to treat neurodegenerative diseases or delay their onset: A perspective, *Neural Regen. Res.* 17 (2022) 1745–1747.
- [28] J.Y. Hwang, K.A. Aromolaran, R.S. Zukin, The emerging field of epigenetics in neurodegeneration and neuroprotection, *Nat. Rev. Neurosci.* 18 (2017) 347–361.
- [29] A. Jowaeid, I. Schmitt, O. Kaut, et al., Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains, *J. Neurosci.* 30 (2010) 6355–6359.
- [30] L. Matsumoto, H. Takuma, A. Tamaoka, et al., CpG demethylation enhances alpha-synuclein expression and affects the pathogenesis of Parkinson's disease, *PLoS One* 5 (2010), e15522.
- [31] F. Albrecht, M. List, C. Bock, et al., DeepBlue epigenomic data server: Programmatic data retrieval and analysis of epigenome region sets, *Nucleic Acids Res.* 44 (2016) W581–W586.
- [32] Z. Xiong, F. Yang, M. Li, et al., EWAS Open Platform: Integrated data, knowledge and toolkit for epigenome-wide association study, *Nucleic Acids Res.* 50 (2022) D1004–D1009.
- [33] Y. Kodama, J. Mashima, T. Kosuge, et al., DDBJ update: The Genomic Expression Archive (GEA) for functional genomics data, *Nucleic Acids Res.* 47 (2019) D69–D73.
- [34] E. Sollis, A. Mosaku, A. Abid, et al., The NHGRI-EBI GWAS Catalog: Knowledge and deposition resource, *Nucleic Acids Res.* 51 (2023) D977–D985.
- [35] D. Bujold, R. Grégoire, D. Brownlee, et al., IHEC data portal. I. Abugessaisa, T. Kasukawa. *Practical Guide to Life Science Databases*, first ed., Springer Nature, Singapore, 2011, pp. 77–94.
- [36] National Centralized Repository for Alzheimer's Disease and Related Dementias (NCRAD). <https://ncrad.iu.edu/>. (Accessed 18 June 2023).
- [37] R.E. Consortium, A. Kundaje, W. Meuleman, et al., Integrative analysis of 111 reference human epigenomes, *Nature* 518 (2015) 317–330.
- [38] T. Barrett, S.E. Wilhite, P. Ledoux, et al., NCBI GEO: Archive for functional genomics data sets: Update, *Nucleic Acids Res.* 41 (2013) D991–D995.
- [39] C. Huttenhower, O. Hofmann, A quick guide to large-scale genomic data mining, *PLoS Comput. Biol.* 6 (2010), e1000779.
- [40] F. Shi, Y. He, Y. Chen, et al., Comparative analysis of multiple neurodegenerative diseases based on advanced epigenetic aging brain, *Front. Genet.* 12



- (2021), 657636.
- [41] S. Mallik, Z. Zhao, Detecting methylation signatures in neurodegenerative disease by density-based clustering of applications with reducing noise, *Sci. Rep.* 10 (2020), 22164.
  - [42] C. Pellegrini, C. Pirazzini, C. Sala, et al., A meta-analysis of brain DNA methylation across sex, age, and Alzheimer's disease points for accelerated epigenetic aging in neurodegeneration, *Front. Aging Neurosci.* 13 (2021), 639428.
  - [43] H.U. Klein, C. McCabe, E. Gjonneska, et al., Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains, *Nat. Neurosci.* 22 (2019) 37–46.
  - [44] P.L. De Jager, Deconstructing the epigenomic architecture of human neurodegeneration, *Neurobiol. Dis.* 153 (2021), 105331.
  - [45] L.F. MacBean, A.R. Smith, K. Lunnon, Exploring beyond the DNA sequence: A review of epigenomic studies of DNA and histone modifications in dementia, *Curr. Genet. Med. Rep.* 8 (2020) 79–92.
  - [46] T. Lu, C.E. Ang, X. Zhuang, Spatially resolved epigenomic profiling of single cells in complex tissues, *Cell* 185 (2022) 4448–4464.e17.
  - [47] Y. Deng, M. Bartosovic, P. Kukanja, et al., Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level, *Science* 375 (2022) 681–686.
  - [48] Y. Deng, M. Bartosovic, S. Ma, et al., Spatial profiling of chromatin accessibility in mouse and human tissues, *Nature* 609 (2022) 375–383.
  - [49] R. Fan, D. Zhang, G. Su, Spatially resolved epigenome-transcriptome co-profiling of mammalian tissues at the cellular level, *Res. Sq.* (2022) 1–26.
  - [50] I.A. Qureshi, M.F. Mehler, Understanding neurological disease mechanisms in the era of epigenetics, *JAMA Neurol* 70 (2013) 703–710.
  - [51] M.S. Rao, T.R. Van Vleet, R. Ciurlionis, et al., Comparison of RNA-seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies, *Front. Genet.* 9 (2018), 636.
  - [52] J. Verheijen, K. Slegers, Understanding Alzheimer disease at the interface between genetics and transcriptomics, *Trends Genet.* 34 (2018) 434–447.
  - [53] S. Han, K. Nho, Y. Lee, Alternative splicing regulation of an Alzheimer's risk variant in *CLU*, *Int. J. Mol. Sci.* 21 (2020), 7079.
  - [54] M. Wang, N.D. Beckmann, P. Roussos, et al., The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease, *Sci. Data* 5 (2018), 180185.
  - [55] D.A. Bennett, A.S. Buchman, P.A. Boyle, et al., Religious orders study and rush memory and aging project, *J. Alzheimers Dis.* 64 (2018) S161–S189.
  - [56] A. Lachmann, D. Torre, A.B. Keenan, et al., Massive mining of publicly available RNA-seq data from human and mouse, *Nat. Commun.* 9 (2018), 1366.
  - [57] Y. Nakamura, Y. Kodama, S. Saruhashi, et al., DDBJ sequence read archive/DDBJ omics archive, *Nat. Preced.* (2010), <https://doi.org/10.1038/npre.2010.5085.1>.
  - [58] SYNAPSE. <https://www.synapse.org/>. (Accessed 18 June 2023).
  - [59] J.I. Satoh, Y. Yamamoto, N. Asahina, et al., RNA-Seq data mining: Down-regulation of *NeuroD6* serves as a possible biomarker for Alzheimer's disease brains, *Dis. Markers* 2014 (2014), 123165.
  - [60] S. Mukherjee, L. Heath, C. Preuss, et al., Molecular estimation of neurodegeneration pseudotime in older brains, *Nat. Commun.* 11 (2020), 5781.
  - [61] K. Hooshmand, G.M. Halliday, S.S. Pineda, et al., Overlap between central and peripheral transcriptomes in Parkinson's disease but not Alzheimer's disease, *Int. J. Mol. Sci.* 23 (2022), 5200.
  - [62] M.B. Hossain, M.K. Islam, A. Adhikari, et al., Bioinformatics approach to identify significant biomarkers, drug targets shared between Parkinson's disease and bipolar disorder: A pilot study, *Bioinform. Biol. Insights* 16 (2022), 11779322221079232.
  - [63] S. Batchu, Progressive multiple sclerosis transcriptome deconvolution indicates increased M2 macrophages in inactive lesions, *Eur. Neurol.* 83 (2020) 433–435.
  - [64] B. Hwang, J.H. Lee, D. Bang, Single-cell RNA sequencing technologies and bioinformatics pipelines, *Exp. Mol. Med.* 50 (2018) 1–14.
  - [65] O.B. Poirion, X. Zhu, T. Ching, et al., Single-cell transcriptomics bioinformatics and computational challenges, *Front. Genet.* 7 (2016), 163.
  - [66] S. Slovin, A. Carissimo, F. Panariello, et al., Single-cell RNA sequencing analysis: A step-by-step overview, *Methods Mol. Biol.* 2284 (2021) 343–365.
  - [67] V. Svensson, E. da Veiga Beltrame, L. Pachter, A curated database reveals trends in single-cell transcriptomics, *Database* 2020 (2020), baaa073.
  - [68] S. Ma, S.B. Lim, Single-cell RNA sequencing in Parkinson's disease, *Bio-medicines* 9 (2021), 368.
  - [69] J. Jiang, C. Wang, R. Qi, et al., scREAD: A single-cell RNA-seq database for Alzheimer's disease, *iScience* 23 (2020), 101769.
  - [70] Q. Wang, S. Ding, Y. Li, et al., The Allen mouse brain common coordinate framework: A 3D reference atlas, *Cell* 181 (2020) 936–953.e20.
  - [71] Y. Hu, S.G. Tattikota, Y. Liu, et al., DRScDB: A single-cell RNA-seq resource for data mining and data comparison across species, *Comput. Struct. Biotechnol. J.* 19 (2021) 2018–2026.
  - [72] P.N. Pushparaj, G. Kalamegam, K.H. Wali Sait, et al., Decoding the role of astrocytes in the entorhinal cortex in Alzheimer's disease using high-dimensional single-nucleus RNA sequencing data and next-generation knowledge discovery methodologies: Focus on drugs and natural product remedies for dementia, *Front. Pharmacol.* 12 (2021), 720170.
  - [73] A.P. Tsai, C. Dong, P.B. Lin, et al., *PLCG2* is associated with the inflammatory response and is induced by amyloid plaques in Alzheimer's disease, *Genome Med.* 14 (2022), 17.
  - [74] A. Grubman, G. Chew, J.F. Ouyang, et al., A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation, *Nat. Neurosci.* 22 (2019) 2087–2097.
  - [75] S.F. Lau, H. Cao, A.K.Y. Fu, et al., Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 25800–25809.
  - [76] H. Mathys, J. Davila-Velderrain, Z. Peng, et al., Single-cell transcriptomic analysis of Alzheimer's disease, *Nature* 570 (2019) 332–337.
  - [77] X. Wang, L. Li, Cell type-specific potential pathogenic genes and functional pathways in Alzheimer's Disease, *BMC Neurol.* 21 (2021), 381.
  - [78] G. Pei, B. Fernandes, Y. Wang, et al., A single-cell atlas of the human brain in Alzheimer's disease and its implications for personalized drug repositioning, *bioRxiv* (2022). <https://doi.org/10.1101/2022.06.14.496100>.
  - [79] M. Wilhelm, J. Schlegl, H. Hahne, et al., Mass-spectrometry-based draft of the human proteome, *Nature* 509 (2014) 582–587.
  - [80] X. Wang, Q. Liu, B. Zhang, Leveraging the complementary nature of RNA-Seq and shotgun proteomics data, *Proteomics* 14 (2014) 2676–2687.
  - [81] T. Maier, M. Güell, L. Serrano, Correlation of mRNA and protein in complex biological samples, *FEBS Lett.* 583 (2009) 3966–3973.
  - [82] E.M. Schoof, B. Furtwängler, N. Üresin, et al., Quantitative single-cell proteomics as a tool to characterize cellular hierarchies, *Nat. Commun.* 12 (2021), 3341.
  - [83] K.M. Verrou, I. Tsamardinos, G. Papoutsoglou, Learning pathway dynamics from single-cell proteomic data: A comparative study, *Cytometry A* 97 (2020) 241–252.
  - [84] V.S. Rao, K. Srinivas, G.N. Sujini, et al., Protein-protein interaction detection: Methods and analysis, *Int. J. Proteom.* 2014 (2014), 147648.
  - [85] A. Zhang, Protein interaction networks: computational analysis, Cambridge University Press, 2009, pp. 1–292.
  - [86] M.W. Gonzalez, M.G. Kann, Chapter 4: Protein interactions and disease, *PLoS Comput. Biol.* 8 (2012), e1002819.
  - [87] R. Craig, J.P. Cortens, R.C. Beavis, Open source system for analyzing, validating, and storing protein identification data, *J. Proteome Res.* 3 (2004) 1234–1242.
  - [88] S. Okuda, Y. Watanabe, Y. Moriya, et al., jPOSTrepo: An international standard data repository for proteomes, *Nucleic Acids Res.* 45 (2017) D1107–D1111.
  - [89] M. Choi, J. Carver, C. Chiva, et al., MassIVE.quant: A community resource of quantitative mass spectrometry-based proteomics datasets, *Nat. Meth.* 17 (2020) 981–984.
  - [90] Proteomic Data Commons. <https://proteomic.datacommons.cancer.gov/pdc/>. (Accessed 18 June 2023).
  - [91] J.A. Vizcaino, E.W. Deutsch, R. Wang, et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination, *Nat. Biotechnol.* 32 (2014) 223–226.
  - [92] L. Martens, H. Hermjakob, P. Jones, et al., PRIDE: The proteomics identifications database, *Proteomics* 5 (2005) 3537–3545.
  - [93] P. Samaras, T. Schmidt, M. Frejno, et al., ProteomicsDB: A multi-omics and multi-organism resource for life science research, *Nucleic Acids Res.* 48 (2020) D1153–D1163.
  - [94] A. Freitas, M. Aroso, S. Rocha, et al., Bioinformatic analysis of the human brain extracellular matrix proteome in neurodegenerative disorders, *Eur. J. Neurosci.* 53 (2021) 4016–4033.
  - [95] S.C. Deolankar, A.H. Patil, D.A.B. Rex, et al., Mapping post-translational modifications in brain regions in Alzheimer's disease using proteomics data mining, *Omics* 25 (2021) 525–536.
  - [96] H. Haytural, R. Benfeitas, S. Schedin-Weiss, et al., Insights into the changes in the proteome of Alzheimer disease elucidated by a meta-analysis, *Sci. Data* 8 (2021), 312.
  - [97] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, et al., Enrichr: A comprehensive gene set enrichment analysis web server 2016 update, *Nucleic Acids Res.* 44 (2016) W90–W97.
  - [98] Y. Kinoshita, T. Uo, S. Jayadev, et al., Potential applications and limitations of proteomics in the study of neurological disease, *Arch. Neurol.* 63 (2006) 1692–1696.
  - [99] A.D. Brunner, M. Thielert, C. Vasilopoulou, et al., Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation, *Mol. Syst. Biol.* 18 (2022), e10798.
  - [100] J.M. Perkel, Single-cell proteomics takes centre stage, *Nature* 597 (2021) 580–582.
  - [101] I. Paul, C. White, I. Turcinovic, et al., Imaging the future: The emerging era of single-cell spatial proteomics, *FEBS J.* 288 (2021) 6990–7001.
  - [102] H. Gallart-Ayala, T. Teav, J. Ivanisevic, Metabolomics meets lipidomics: Assessing the small molecule component of metabolism, *BioEssays* 42 (2020), e2000052.
  - [103] R. Wang, B. Li, S.M. Lam, et al., Integration of lipidomics and metabolomics for in-depth understanding of cellular mechanism and disease progression, *Yi Chuan Xue Bao* 47 (2020) 69–83.
  - [104] M.A. Alves, S. Lamichhane, A. Dickens, et al., Systems biology approaches to study lipidomics in health and disease, *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids.* 1866 (2021), 158857.
  - [105] D.B. Castellanos, C.A. Martín-Jiménez, F. Rojas-Rodríguez, et al., Brain lipidomics as a rising field in neurodegenerative contexts: Perspectives with Machine Learning approaches, *Front. Neuroendocrinol.* 61 (2021), 100899.



- [106] B.B. Misra, C. Langefeld, M. Olivier, et al., Integrated omics: Tools, advances and future approaches, *J. Mol. Endocrinol.* 62 (2019) R21–R45.
- [107] A. Schumacher-Schuh, A. Bieger, W.V. Borelli, et al., Advances in proteomic and metabolomic profiling of neurodegenerative diseases, *Front. Neurol.* 12 (2021), 792227.
- [108] Y. Shao, W. Le, Recent advances and perspectives of metabolomics-based investigations in Parkinson's disease, *Mol. Neurodegener.* 14 (2019), 3.
- [109] P. Reveglia, C. Paolillo, G. Ferretti, et al., Challenges in LC-MS-based metabolomics for Alzheimer's disease early detection: Targeted approaches versus untargeted approaches, *Metabolomics* 17 (2021), 78.
- [110] R.C. Petersen, P.S. Aisen, L.A. Beckett, et al., Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization, *Neurology* 74 (2010) 201–209.
- [111] D.S. Wishart, M.J. Lewis, J.A. Morrissey, et al., The human cerebrospinal fluid metabolome, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 871 (2008) 164–173.
- [112] D.S. Wishart, D. Tzur, C. Knox, et al., HMDB: The human metabolome database, *Nucleic Acids Res.* 35 (2007) D521–D526.
- [113] D.S. Wishart, C. Knox, A.C. Guo, et al., HMDB: A knowledgebase for the human metabolome, *Nucleic Acids Res.* 37 (2009) D603–D610.
- [114] D.S. Wishart, T. Jewison, A.C. Guo, et al., HMDB 3.0-The human metabolome database in 2013, *Nucleic Acids Res.* 41 (2013) D801–D807.
- [115] D.S. Wishart, Y.D. Feunang, A. Marcu, et al., HMDB 4.0: The human metabolome database for 2018, *Nucleic Acids Res.* 46 (2018) D608–D617.
- [116] D.S. Wishart, A.C. Guo, E. Oler, et al., HMDB 5.0: The human metabolome database for 2022, *Nucleic Acids Res.* 50 (2022) D622–D631.
- [117] K. Watanabe, E. Yasugi, M. Oshima, How to search the glycolipid data in "LIPIDBANK for Web" the newly developed lipid database in Japan, *Trends Glycosci. Glycotechnol.* 12 (2000) 175–184.
- [118] T. Kind, K.H. Liu, D.Y. Lee, et al., LipidBlast in silico tandem mass spectrometry database for lipid identification, *Nat. Meth.* 10 (2013) 755–758.
- [119] D. Cotter, A. Maer, C. Guda, et al., LMPD: LIPID MAPS proteome database, *Nucleic Acids Res.* 34 (2006) D507–D510.
- [120] T.B. Mracica, A. Anghel, C.F. Ion, et al., MetaboAge DB: A repository of known ageing-related changes in the human metabolome, *Biogerontology* 21 (2020) 763–771.
- [121] K. Haug, R.M. Salek, P. Conesa, et al., MetaboLights: An open-access general-purpose repository for metabolomics studies and associated meta-data, *Nucleic Acids Res.* 41 (2013) D781–D786.
- [122] M. Sud, E. Fahy, D. Cotter, et al., Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools, *Nucleic Acids Res.* 44 (2016) D463–D470.
- [123] MetabolomeXchange. <http://www.metabolomexchange.org/site/>. (Accessed 18 June 2023).
- [124] N. Psychogios, D.D. Hau, J. Peng, et al., The human serum metabolome, *PLoS One* 6 (2011), e16957.
- [125] D.K. Barupal, R. Baillie, S. Fan, et al., Sets of coregulated serum lipids are associated with Alzheimer's disease pathophysiology, *Alzheimers Dement.* (Amst) 11 (2019) 619–627.
- [126] Y. Tang, S. Shah, K.S. Cho, et al., Metabolomics in primary open angle glaucoma: A systematic review and meta-analysis, *Front. Neurosci.* 16 (2022), 835736.
- [127] Y. Perez-Riverol, M. Bai, F. da Veiga Leprevost, et al., Discovering and linking public omics data sets using the Omics Discovery Index, *Nat. Biotechnol.* 35 (2017) 406–409.
- [128] J. Pu, Y. Yu, Y. Liu, et al., MENDA: A comprehensive curated resource of metabolic characterization in depression, *Brief Bioinform* 21 (2020) 1455–1464.
- [129] X. Han, R. Wang, Y. Zhou, et al., Mapping the mouse cell atlas by microwell-seq, *Cell* 172 (2018) 1091–1107.e17.
- [130] The Tabula Muris Consortium, Overall coordination, Logistical coordination, et al., Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris, *Nature* 562 (2018) 367–372.
- [131] J. Liao, X. Lu, X. Shao, et al., Uncovering an organ's molecular architecture at single-cell resolution by spatially resolved transcriptomics, *Trends Biotechnol* 39 (2021) 43–58.
- [132] J. Cheng, J. Liao, X. Shao, et al., Multiplexing methods for simultaneous large-scale transcriptomic profiling of samples at single-cell resolution, *Adv. Sci.* (Weinh) 8 (2021), e2101229.
- [133] K.H. Chen, A.N. Boettiger, J.R. Moffitt, et al., RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells, *Science* 348 (2015), aaa6090.
- [134] P.L. Ståhl, F. Salmén, S. Vickovic, et al., Visualization and analysis of gene expression in tissue sections by spatial transcriptomics, *Science* 353 (2016) 78–82.
- [135] S.G. Rodrigues, R.R. Stickels, A. Goeva, et al., Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution, *Science* 363 (2019) 1463–1467.
- [136] C.L. Eng, M. Lawson, Q. Zhu, et al., Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH, *Nature* 568 (2019) 235–239.
- [137] R. Satija, J.A. Farrell, D. Gennert, et al., Spatial reconstruction of single-cell gene expression data, *Nat. Biotechnol.* 33 (2015) 495–502.
- [138] K. Achim, J.B. Pettit, L.R. Saraiva, et al., High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin, *Nat. Biotechnol.* 33 (2015) 503–509.
- [139] D. Franjic, M. Skarica, S. Ma, et al., Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells, *Neuron* 110 (2022) 452–469.e14.
- [140] A.M. Femino, F.S. Fay, K. Fogarty, et al., Visualization of single RNA transcripts in situ, *Science* 280 (1998) 585–590.
- [141] C. Giesen, H.A.O. Wang, D. Schapiro, et al., Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry, *Nat. Meth.* 11 (2014) 417–422.
- [142] G. Gut, M.D. Herrmann, L. Pelkmans, Multiplexed protein maps link subcellular organization to cellular states, *Science* 361 (2018), eaar7042.
- [143] W. Chen, A. Lu, K. Craessaerts, et al., Spatial transcriptomics and *in situ* sequencing to study Alzheimer's disease, *Cell* 182 (2020) 976–991.e19.
- [144] S. Prokop, K.R. Miller, S.R. Labra, et al., Impact of TREM2 risk variants on brain region-specific immune activation and plaque microenvironment in Alzheimer's disease patient brain samples, *Acta Neuropathol.* 138 (2019) 613–630.
- [145] J.F. Navarro, D.L. Croteau, A. Jurek, et al., Spatial transcriptomics reveals genes associated with dysregulated mitochondrial functions and stress signaling in Alzheimer disease, *iScience* 23 (2020), 101556.
- [146] J. Aguila, S. Cheng, N. Kee, et al., Spatial RNA sequencing identifies robust markers of vulnerable and resistant human midbrain dopamine neurons and their expression in Parkinson's disease, *Front. Mol. Neurosci.* 14 (2021), 699562.
- [147] T. Kamath, A. Abdullaouf, S.J. Burrell, et al., Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease, *Nat. Neurosci.* 25 (2022) 588–595.
- [148] S. Smajić, C.A. Prada-Medina, Z. Landoulsi, et al., Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state, *Brain* 145 (2022) 964–978.
- [149] C.H. Lo, M. Skarica, M. Mansoor, et al., Astrocyte heterogeneity in multiple sclerosis: Current understanding and technical challenges, *Front. Cell. Neurosci.* 15 (2021), 726479.
- [150] L. Schirmer, D. Velmeshev, S. Holmqvist, et al., Neuronal vulnerability and multilineage diversity in multiple sclerosis, *Nature* 573 (2019) 75–82.
- [151] M. Kaufmann, H. Evans, A.L. Schaupp, et al., Identifying CNS-colonizing T cells as potential therapeutic targets to prevent progression of multiple sclerosis, *Med* 2 (2021) 296–312.e8.
- [152] M. Absinta, D. Maric, M. Gharagzloo, et al., A lymphocyte-microglia-astrocyte axis in chronic active multiple sclerosis, *Nature* 597 (2021) 709–714.
- [153] M. Kaufmann, A.L. Schaupp, R. Sun, et al., Identification of early neurodegenerative pathways in progressive multiple sclerosis, *Nat. Neurosci.* 25 (2022) 944–955.
- [154] B. Pardo, A. Spangler, L.M. Weber, et al., spatialLIBD: An R/Bioconductor package to visualize spatially-resolved transcriptomics data, *BMC Genomics* 23 (2022), 434.
- [155] D. Righelli, L.M. Weber, H.L. Crowell, et al., SpatialExperiment: Infrastructure for spatially-resolved transcriptomics data in R using Bioconductor, *Bioinformatics* 38 (2022) 3128–3131.
- [156] C.K. Mah, N. Ahmed, N. Lopez, et al., Bento: A toolkit for subcellular analysis of spatial transcriptomics data, *bioRxiv* (2023). <https://doi.org/10.1101/2022.06.10.495510>.
- [157] L.M. Breckels, C.M. Mulvey, K.S. Lilley, et al., A Bioconductor workflow for processing and analysing spatial proteomics data, *F1000Research* 5 (2016), 2926.
- [158] G. Palla, H. Spitzer, M. Klein, et al., Squidpy: A scalable framework for spatial omics analysis, *Nat. Meth.* 19 (2022) 171–178.
- [159] S. Vickovic, B. Lötstedt, J. Klughammer, et al., SM-Omics is an automated platform for high-throughput spatial multi-omics, *Nat. Commun.* 13 (2022), 795.
- [160] A. Martinelli, M. Rapsomaniki, ATHENA: Analysis of tumor heterogeneity from spatial omics measurements, *Bioinformatics* 38 (2022) 3151–3153.
- [161] M.A. Kennedy, W.A. Hofstadter, I.M. Cristea, TRANSPIRE: A computational pipeline to elucidate intracellular protein movements from spatial proteomics data sets, *J. Am. Soc. Mass Spectrom.* 31 (2020) 1422–1439.
- [162] J.A. Christopher, A. Geladaki, C.S. Dawson, et al., Subcellular transcriptomics and proteomics: A comparative methods review, *Mol. Cell. Proteom.* 21 (2022), 100186.
- [163] A. Regev, S.A. Teichmann, E.S. Lander, et al., The human cell atlas, *eLife* 6 (2017), e27041.
- [164] HuBMAP Consortium, The human body at cellular resolution: The NIH Human Biomolecular Atlas Program, *Nature* 574 (2019) 187–192.
- [165] M. Frenkel-Morgenstern, A.A. Cohen, N. Geva-Zatorsky, et al., Dynamic proteomics: A database for dynamics and localizations of endogenous fluorescently-tagged proteins in living human cells, *Nucleic Acids Res.* 38 (2010) D508–D512.
- [166] R. Dries, Q. Zhu, R. Dong, et al., Giotto: A toolbox for integrative analysis and visualization of spatial expression data, *Genome Biol.* 22 (2021), 78.
- [167] Z. Xu, W. Wang, T. Yang, et al., STOmicsDB: A database of Spatial Transcriptomic data, *bioRxiv* (2022), <https://doi.org/10.1101/2022.03.11.481421>.
- [168] Z. Fan, R. Chen, X. Chen, SpatialDB: A database for spatially resolved transcriptomes, *Nucleic Acids Res.* 48 (2020) D233–D237.
- [169] J. Qian, J. Liao, Z. Liu, et al., Reconstruction of the cell pseudo-space from single-cell RNA sequencing data with scSpace, *Nat. Commun.* 14 (2023), 2484.

- [170] J. Liao, J. Qian, Y. Fang, et al., De novo analysis of bulk RNA-seq data at spatially resolved single-cell resolution, *Nat. Commun.* 13 (2022), 6498.
- [171] A. Zeisel, A.B. Muñoz-Manchado, S. Codeluppi, et al., Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq, *Science* 347 (2015) 1138–1142.
- [172] F. Bao, Y. Deng, S. Wan, et al., Integrative spatial analysis of cell morphologies and transcriptional states with MUSE, *Nat. Biotechnol.* 40 (2022) 1200–1209.
- [173] V. Marx, Method of the year: Spatially resolved transcriptomics, *Nat. Meth.* 18 (2021) 9–14.
- [174] S.K. Longo, M.G. Guo, A.L. Ji, et al., Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics, *Nat. Rev. Genet.* 22 (2021) 627–644.
- [175] X. Shao, X. Lu, J. Liao, et al., New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data, *Protein Cell* 11 (2020) 866–880.
- [176] X. Shao, J. Liao, C. Li, et al., CellTalkDB: A manually curated database of ligand-receptor interactions in humans and mice, *Brief. Bioinform.* 22 (2021), bbaa269.
- [177] X. Shao, C. Li, H. Yang, et al., Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk, *Nat. Commun.* 13 (2022), 4429.
- [178] Y. Fangma, M. Liu, J. Liao, et al., Dissecting the brain with spatially resolved multi-omics, *J. Pharm. Anal.* (2023), <https://doi.org/10.1016/j.jpaha.2023.04.003>.
- [179] T.J. Sejnowski, P.S. Churchland, J.A. Movshon, Putting big data to good use in neuroscience, *Nat. Neurosci.* 17 (2014) 1440–1441.
- [180] H. Lee, J.J. Lee, N.Y. Park, et al., Multi-omic analysis of selectively vulnerable motor neuron subtypes implicates altered lipid metabolism in ALS, *Nat. Neurosci.* 24 (2021) 1673–1685.
- [181] L. Xicota, F. Ichou, F.X. Lejeune, et al., Multi-omics signature of brain amyloid deposition in asymptomatic individuals at-risk for Alzheimer's disease: The INSIGHT-preAD study, *EBioMedicine* 47 (2019) 518–528.
- [182] E. Puris, S. Kouřil, L. Najdekr, et al., Metabolomic, lipidomic and proteomic characterisation of lipopolysaccharide-induced inflammation mouse model, *Neuroscience* 496 (2022) 165–178.
- [183] C. Clark, L. Dayon, M. Masoodi, et al., An integrative, hypothesis-free, multi-omics approach uncovers biological pathway alterations in Alzheimer's disease, *Alzheimers. Dement.* 16 (2020), e038563.
- [184] S. Lee, N.A. Devanney, L.R. Golden, et al., APOE modulates microglial immunometabolism in response to age, amyloid pathology, and inflammatory challenge, *Cell Rep.* 42 (2023), 112196.
- [185] M.B. O'Rourke, S.E.L. Town, P.V. Dalla, et al., What is normalization? the strategies employed in top-down and bottom-up proteome analysis workflows, *Proteomes* 7 (2019), 29.
- [186] A. Conesa, P. Madrigal, S. Tarazona, et al., A survey of best practices for RNA-seq data analysis, *Genome Biol.* 17 (2016), 13.
- [187] P. Yang, H. Huang, C. Liu, Feature selection revisited in the single-cell era, *Genome Biol.* 22 (2021), 321.
- [188] Á. Torres-Martos, M. Bustos-Aibar, A. Ramírez-Mena, et al., Omics data preprocessing for machine learning: A case study in childhood obesity, *Genes* 14 (2023), 248.
- [189] S. Nataf, M. Guillen, L. Pays, TGFβ1-mediated gliosis in multiple sclerosis spinal cords is favored by the regionalized expression of HOXA5 and the age-dependent decline in androgen receptor ligands, *Int. J. Mol. Sci.* 20 (2019), 5934.
- [190] M.E. Garcia-Segura, B.R. Durainayagam, S. Liggi, et al., Pathway-based integration of multi-omics data reveals lipidomics alterations validated in an Alzheimer's disease mouse model and risk loci carriers, *J. Neurochem.* 164 (2023) 57–76.
- [191] G. Zhou, S. Li, J. Xia, Network-based approaches for multi-omics integration, *Methods Mol. Biol.* 2104 (2020) 469–487.
- [192] M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (2000) 27–30.
- [193] M. Kanehisa, M. Furumichi, Y. Sato, et al., KEGG for taxonomy-based analysis of pathways and genomes, *Nucleic Acids Res.* 51 (2023) D587–D592.
- [194] M. Ashburner, C.A. Ball, J.A. Blake, et al., Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.* 25 (2000) 25–29.
- [195] Gene Ontology Consortium, The Gene Ontology resource: Enriching a Gold mine, *Nucleic Acids Res.* 49 (2021) D325–D334.
- [196] G. Dennis Jr., B.T. Sherman, D.A. Hosack, et al., DAVID: Database for annotation, visualization, and integrated discovery, *Genome Biol.* 4 (2003), P3.
- [197] P.D. Thomas, D. Ebert, A. Muruganujan, et al., PANTHER: Making genome-scale phylogenetics accessible to all, *Protein Sci.* 31 (2022) 8–22.
- [198] H. Mi, A. Muruganujan, X. Huang, et al., Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0), *Nat. Protoc.* 14 (2019) 703–721.
- [199] A. Subramanian, P. Tamayo, V.K. Mootha, et al., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 15545–15550.
- [200] A. Krämer, J. Green, J. Pollard Jr., et al., Causal analysis approaches in Ingenuity Pathway Analysis, *Bioinformatics* 30 (2014) 523–530.
- [201] D. Szklarczyk, A.L. Gable, D. Lyon, et al., STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Res.* 47 (2019) D607–D613.
- [202] P. Shannon, A. Markiel, O. Ozier, et al., Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504.
- [203] F. Martin, T.M. Thomson, A. Sewer, et al., Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks, *BMC Syst. Biol.* 6 (2012), 54.
- [204] A.L. Tarca, S. Draghici, P. Khatri, et al., A novel signaling pathway impact analysis, *Bioinformatics* 25 (2009) 75–82.
- [205] L. Cottret, C. Frainay, M. Chazalviel, et al., MetExplore: Collaborative edition and exploration of metabolic networks, *Nucleic Acids Res.* 46 (2018) W495–W502.
- [206] T.C. Kuo, T.F. Tian, Y.J. Tseng, 3Omics: A web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data, *BMC Syst. Biol.* 7 (2013), 64.
- [207] T. Liu, P. Salguero, M. Petek, et al., PaintOmics 4: New tools for the integrative analysis of multi-omics datasets supported by multiple pathway databases, *Nucleic Acids Res.* 50 (2022) W551–W559.
- [208] I. Subramanian, S. Verma, S. Kumar, et al., Multi-omics data integration, interpretation, and its application, *Bioinform. Biol. Insights* 14 (2020), 1177932219899051.
- [209] Y. Xu, R.P. McCord, Diagonal integration of multimodal single-cell data: Potential pitfalls and paths forward, *Nat. Commun.* 13 (2022), 3505.
- [210] C. Stark, B.J. Breitkreutz, T. Reguly, et al., BioGRID: A general repository for interaction datasets, *Nucleic Acids Res.* 34 (2006) D535–D539.
- [211] H. Fanaee-T, M. Thoresen, Multi-insight visualization of multi-omics data via ensemble dimension reduction and tensor factorization, *Bioinformatics* 35 (2019) 1625–1633.
- [212] N. Vahabi, G. Michailidis, Unsupervised multi-omics data integration methods: A comprehensive review, *Front. Genet.* 13 (2022), 854752.
- [213] D.D. Lee, H.S. Seung, Algorithms for Non-negative Matrix Factorization, *Proceedings of Adv. Neural Inf. Process.*, Dec 3 – 8, 2001, Vancouver, Canada, 2001.
- [214] M. Pierre-Jean, J.F. Deleuze, E. le Floch, et al., Clustering and variable selection evaluation of 13 unsupervised methods for multi-omics data integration, *Brief. Bioinform.* 21 (2020) 2011–2030.
- [215] Z. Yang, G. Michailidis, A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data, *Bioinformatics* 32 (2016) 1–8.
- [216] S. MacEachern, P. Müller, Efficient MCMC Schemes for Robust Model Extensions Using Encompassing Dirichlet Process Mixture Models - Robust Bayesian Analysis, P. Diggle, S. Zeger (Eds.), *Lecture Notes in Statistics*, first ed., Springer, New York, 2000, pp. 295–315.
- [217] L. Cowen, T. Ideker, B.J. Raphael, et al., Network propagation: A universal amplifier of genetic associations, *Nat. Rev. Genet.* 18 (2017) 551–562.
- [218] W. Ye, G. Ji, P. Ye, et al., scNPF: An integrative framework assisted by network propagation and network fusion for preprocessing of single-cell RNA-seq data, *BMC Genomics* 20 (2019), 347.
- [219] P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis, *BMC Bioinformatics* 9 (2008), 559.
- [220] S. Dolédec, D. Chessel, Co-inertia analysis: an alternative method for studying species-environment relationships, *Freshw. Biol.* 31 (1994) 277–294.
- [221] K. Sankaran, S.P. Holmes, Multitable methods for microbiome data integration, *Front. Genet.* 10 (2019), 627.
- [222] R. Shen, A.B. Olshen, M. Ladanyi, Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis, *Bioinformatics* 25 (2009) 2906–2912.
- [223] Q. Mo, S. Wang, V.E. Seshan, et al., Pattern discovery and cancer gene identification in integrated cancer genomic data, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 4245–4250.
- [224] E.F. Lock, K.A. Hoadley, J.S. Marron, et al., Joint and individual variation explained (jive) for integrated analysis of multiple data types, *Ann. Appl. Stat.* 7 (2013) 523–542.
- [225] R. Louhimo, S. Hautaniemi, CNAMet: An R package for integrating copy number, methylation and expression data, *Bioinformatics* 27 (2011) 887–888.
- [226] C.J. Vaske, S.C. Benz, J.Z. Sanborn, et al., Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM, *Bioinformatics* 26 (2010) i237–i245.
- [227] J.-C. Park, N. Barahona-Torres, S.-Y. Jang, et al., Multi-omics-based auto-phagy-related untypical subtypes in patients with cerebral amyloid pathology, *Adv. Sci. (Weinh.)* 9 (2022), e2201212.
- [228] A. Catanese, S. Rajkumar, D. Sommer, et al., Multiomics and machine-learning identify novel transcriptional and mutational signatures in amyotrophic lateral sclerosis, *Brain* (2023), awad075.
- [229] S. Lovestone, P. Francis, I. Kloszewska, et al., AddNeuroMed: The European collaboration for the discovery of novel biomarkers for Alzheimer's disease, *Ann. N. Y. Acad. Sci.* 1180 (2009) 36–46.
- [230] A. Hye, S. Lynham, M. Thambisetty, et al., Proteome-based plasma biomarkers for Alzheimer's disease, *Brain* 129 (2006) 3042–3050.
- [231] J. Xu, G. Bankov, M. Kim, et al., Integrated lipidomics and proteomics network analysis highlights lipid and immunity pathways associated with Alzheimer's disease, *Transl. Neurodegener.* 9 (2020), 36.
- [232] N.T. Seyfried, E.B. Dammer, V. Swarup, et al., A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease, *Cell Syst* 4 (2017) 60–72.e4.

- [233] J.B. Toledo, M. Arnold, G. Kastenmüller, et al., Metabolic network failures in Alzheimer's disease: A biochemical roadmap, *Alzheimers. Dement.* 13 (2017) 965–984.
- [234] E. Bonnet, L. Calzone, T. Michoel, Integrative multi-omics module network inference with Lemon-Tree, *PLoS Comput. Biol.* 11 (2015), e1003983.
- [235] S. Jamal, S. Goyal, A. Shanker, et al., Integrating network, sequence and functional features using machine learning approaches towards identification of novel Alzheimer genes, *BMC Genomics* 17 (2016), 807.
- [236] J. Menche, E. Guney, A. Sharma, et al., Integrating personalized gene expression profiles into predictive disease-associated gene pools, *NPJ Syst. Biol. Appl.* 3 (2017), 10.
- [237] P.T. Ram, J. Mendelsohn, G.B. Mills, Bioinformatics and systems biology, *Mol. Oncol.* 6 (2012) 147–154.
- [238] N.S. Buchan, D.K. Rajpal, Y. Webster, et al., The role of translational bioinformatics in drug discovery, *Drug Discov. Today* 16 (2011) 426–434.
- [239] J. Leipzig, A review of bioinformatic pipeline frameworks, *Brief Bioinform.* 18 (2017) 530–536.
- [240] A. Gupta, S. Gupta, S.K. Jatawa, et al., A simplest bioinformatics pipeline for whole transcriptome sequencing: Overview of the processing and steps from raw data to downstream analysis, *bioRxiv* (2019), <https://doi.org/10.1101/836973>.
- [241] L. Siegwald, H. Touzet, Y. Lemoine, et al., Assessment of Common and Emerging Bioinformatics Pipelines for Targeted Metagenomics, *PLoS One* 12 (2017), e0169563.
- [242] C.P. Moritz, T. Mühlhaus, S. Tenzer, et al., Poor transcript-protein correlation in the brain: Negatively correlating gene products reveal neuronal polarity as a potential cause, *J. Neurochem.* 149 (2019) 582–604.
- [243] M. Jafari, Y. Guan, D.C. Wedge, et al., Re-evaluating experimental validation in the Big Data Era: a conceptual argument, *Genome Biol.* 22 (2021), 71.
- [244] C. Manzoni, D.A. Kia, J. Vandrovцова, et al., Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences, *Brief Bioinform.* 19 (2018) 286–302.
- [245] C.H. Lo, J. Zeng, Defective lysosomal acidification: A new prognostic marker and therapeutic target for neurodegenerative diseases, *Transl. Neurodegener.* 12 (2023), 29.
- [246] D. Pitt, C.H. Lo, S.A. Gauthier, et al., Toward precision phenotyping of multiple sclerosis, *Neurol. Neuroimmunol. Neuroinflamm.* 9 (2022), e200025.
- [247] K.S. Suh, S. Sarojini, M. Youssif, et al., Tissue banking, bioinformatics, and electronic medical records: The front-end requirements for personalized medicine, *J. Oncol.* 2013 (2013), 368751.
- [248] A. Kolodkin, E. Simeonidis, R. Balling, et al., Understanding complexity in neurodegenerative diseases: in silico reconstruction of emergence, *Front. Physiol.* 3 (2012), 291.
- [249] S. Golriz Khatami, S. Mubeen, M. Hofmann-Apitius, Data science in neurodegenerative disease: Its capabilities, limitations, and perspectives, *Curr. Opin. Neurol.* 33 (2020) 249–254.
- [250] M.A. Myszczyńska, P.N. Ojamies, A.M.B. Lacoste, et al., Applications of machine learning to diagnosis and treatment of neurodegenerative diseases, *Nat. Rev. Neurol.* 16 (2020) 440–456.
- [251] A. Mammoliti, P. Smirnov, M. Nakano, et al., Orchestrating and sharing large multimodal data for transparent and reproducible research, *Nat. Commun.* 12 (2021), 5797.
- [252] S. Lam, A. Bayraktar, C. Zhang, et al., A systems biology approach for studying neurodegenerative diseases, *Drug Discov. Today* 25 (2020) 1146–1159.