TUM School of Life Sciences

TECHNISCHE UNIVERSITÄT MÜNCHEN

A network-based, multi-omics integration framework for Alzheimer's disease

Maria Anna Ulmer

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr. Dr. Fabian J. Theis

Prüfer der Dissertation:

1. Prof. Dr. Jan Baumbach
2. Prof. Karsten Suhre, Ph.D.

Die Dissertation wurde am 27.01.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 31.08.2023 angenommen.

HELMHOLTZ MUNICH

TUM School of Life Sciences

A network-based, multi-omics integration framework for Alzheimer's disease

Maria Anna Ulmer

(née Wörheide)

2023

Acknowledgments

The past years have been a wild ride and there are a lot of people who have been a part of this journey to whom I would like to express my gratitude.

I would like to start by thanking my amazing supervisor Dr. Matthias Arnold. Thank you for making this project possible and providing me with such excellent supervision. I am really grateful for your guidance, ideas, and support throughout this – at times – intimidating endeavor. You were always available when I needed help, stayed in the office until midnight to submit papers, and provided me with great advice - thank you!

A special thank you also to Dr. Gabi Kastenmüller for giving me the opportunity to join your lovely group, getting me started, and for your constant support, helpful feedback, and encouraging words.

Furthermore, I want to thank my examination committee Prof. Dr. Jan Baumbach and Prof. Dr. Karsten Suhre, as well as Dr. Josch Pauling for taking the time to provide me with valuable feedback and guidance.

Thank you also to my wonderful SysMet colleagues, without you all this experience would not have been the same! A special thank you to Patrick, Nick, and Tim, who welcomed me to the group with open arms and - most importantly - always had snacks. Thank you also to Johannes, Werner, Dani, Tong, Bharadwaj, Orhan, and Leonard for the fun hiking adventures and office chats.

I had the pleasure of working with lovely collaborators and would also like to use this chance to thank all the study participants who made this work possible.

Lastly, I would like to thank the most important people in my life; Connie, Gert, Jack, Max, Elke, Martin, Lisa, and my amazing friends. Words cannot express how grateful I am to have such an amazing support system - you guys rock.

Mum and Dad - thank you for providing me with an abundance of love and opportunities.

Max - I could not have done this without you. Thank you for always believing in me.

Abstract

Advances in high-throughput technologies have provided us with an unprecedented abundance of molecular datasets. For diseases with complex and multi-factorial etiology, such as Alzheimer's disease (AD), the integrative analysis of different omics data modalities is the key to understanding molecular pathomechanisms underlying the disease and translation into effective, novel therapeutic approaches. However, embedding single-omics disease associations into the wider context of multi-level molecular changes remains a central challenge. In this thesis, I provide an overview of existing multi-omics integration strategies and develop a novel network-based integration framework that I apply to create the AD Atlas, a publicly available web-based multi-omics resource for Alzheimer's research.

In the first part of this work, I provide an overview of large-scale, multi-omics integration strategies, focusing on methods that enable the integration of metabolomics data with other omics datasets. I describe aspects of a typical integration workflow, highlighting major challenges, such as heterogeneous data modalities, differing identifier namespaces, and disjoint sample sets. I distinguish between (i) knowledge-based integration, i.e., methods that use external information to connect different biological entities, (ii) data-driven integration, i.e., methods that use statistical and machine learning techniques to infer interconnections from correlation structures within the data, and (iii) composite networks, i.e., methods that represent relationships between omics layers in heterogeneous network structures and can be built in a knowledge-based, data-driven or hybrid manner. Furthermore, I introduce AD, a devastating neurodegenerative disorder characterized by the accumulation of pathogenic proteins in the brain causing neurodegeneration and cognitive impairment. There is no curative treatment for AD, and previous studies have implicated complex molecular changes across various omics layers in disease onset and progression.

In the second part, I introduce the AD Atlas (www.adatlas.org), an online multi-omics resource that integrates over 25 large studies providing disease-relevant information on 20,363 protein-coding genes, 8,396 proteins, 1,328 metabolites, and 43 AD-related phenotypes. I describe the methods, tools, and considerations that were used to create the resource, including a novel network-based integration framework. I used this to integrate results from numerous AD-specific omics studies from the Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD), NIA Genetics of Alzheimer's Disease Data Storage

Site (NIAGADS), and other interdisciplinary initiatives into a comprehensive network structure and complement them with molecular associations from large-scale population-based studies to provide a global view of molecular changes observed in AD. To assess the biological content of the AD Atlas, I used deep learning-based dimensionality reduction and AD-related functional domain annotations, finding evidence that the network structure as a whole captures relevant biological information.

Finally, I describe the publicly available user interface I implemented to enable efficient access to the AD Atlas data. This network-based platform allows researchers to formulate molecular hypotheses and retrieve clinically relevant findings that can be validated in follow-up analyses or experiments. Using multiple showcases, I demonstrate the utility and relevance of this resource for contextualization of AD research results and subsequent downstream analyses. For example, using two known AD risk genes as input, I show how the AD Atlas can be used to identify drug repositioning candidates. This approach identifies Candesartan and Levetiracetam, two compounds currently tested in clinical trials (reproducing previous research), and multiple other candidates that may be equally promising (adding additional insights).

In conclusion, the presented AD Atlas is a unique multi-omics resource that provides global molecular views on AD as well as a large set of analysis tools to contextualize molecular hypotheses in a multi-omics context. The work and future extensions presented in this thesis are broadly applicable and thus provide a global framework for the integrative analysis of complex diseases.

Zusammenfassung

Durch die Weiterentwicklung und stetige Verbesserung von modernen Hochdurchsatztechnologien werden immer mehr qualitativ hochwertige molekulare '-Omics' Datensätze generiert. Bei Krankheiten mit komplexer und multifaktorieller Ätiologie, wie z.B. Morbus Alzheimer (AD), ist die integrative Analyse dieser Omics-Daten ein wertvoller Ansatz um die der Krankheit zugrunde liegenden Pathomechanismen zu verstehen und neue therapeutische Ansätze zu entwickeln. Die Betrachtung einzelner Erkenntnisse im Kontext vielschichtiger molekularer Veränderungen bleibt jedoch eine zentrale Herausforderung. In dieser Arbeit stelle ich verschiedene Ansätze vor um heterogene molekulare Daten (Multi-Omics) zu integrieren. Zudem entwickle ich eine neue netzwerkbasierte Integrationsstrategie, die ich nutze um den AD Atlas zu entwickeln, eine öffentlich und frei zugängliche Multi-Omics-Ressource für die Alzheimer-Forschung.

Im ersten Teil dieser Arbeit gebe ich einen Überblick über globale Multi-Omics-Integrationsstrategien und konzentriere mich dabei auf Methoden, die die Integration von Metabolomics-Daten mit anderen Omics-Datensätzen ermöglichen. Ich beschreibe die unterschiedlichen Aspekte eines typischen Integrations-Workflows und gehe auf wichtige Herausforderungen ein, wie z.B. heterogene Datenmodalitäten, unterschiedliche Bezeichnungen für biologische Entitäten (z.B. Gennamen) und disjunkte Probensätze. Ich unterscheide zwischen (i) wissensbasierter Integration, d.h. Methoden, die externe Informationen nutzen, um verschiedene biologische Entitäten miteinander zu verbinden, (ii) datengesteuerter Integration, d.h. Methoden, die statistische Methoden nutzen, um aus Korrelationsstrukturen innerhalb der Daten auf Zusammenhänge zu schließen, und (iii) zusammengesetzten Netzwerken, d.h. Methoden, die Beziehungen zwischen den einzelnen Omics-Daten in heterogenen Netzwerkstrukturen darstellen und auf wissensbasierte, datengesteuerte, oder hybride Weise aufgebaut werden. Des Weiteren stelle ich Morbus Alzheimer vor, eine verheerende neurodegenerative Erkrankung, die durch die Ansammlung pathogener Proteine im Gehirn gekennzeichnet ist und zu Neurodegeneration und kognitiven Beeinträchtigungen führt. Für Alzheimer gibt es noch keine Heilung und Studien zeigen, dass die Entstehung und der Verlauf dieser Krankheit durch komplexe molekulare Veränderungen beeinflusst werden.

Im zweiten Teil stelle ich den AD Atlas (www.adatlas.org) vor, eine Online-Multiomics-Ressource, die mehr als 25 große wissenschaftliche Studien integriert und krankheitsrelevante Informationen für über 20.363 proteinkodierende Gene, 8.396 Proteine, 1.328 Metabolite und 43 Alzheimer-spezifische Phänotypen beinhaltet. Ich beschreibe die Methoden und Werkzeuge, die zur Erstellung der Ressource verwendet wurden, einschließlich eines neuen netzwerkbasierten Integrationskonzepts. Mit diesem Ansatz habe ich die Ergebnisse zahlreicher quantitativer Omics-Studien aus Alzheimer Kohorten und interdisziplinären Initiativen in eine umfassende Netzwerkstruktur integriert und mit molekularen Assoziationen aus bevölkerungsbasierten Studien ergänzt, um eine globale Sicht auf die molekularen Veränderungen in Alzheimer zu bekommen. Für die Bewertung des biologischen Inhalts des AD Atlas, habe ich eine auf Deep Learning basierende Dimensionalitätsreduktion und Alzheimer-bezogene funktionelle Annotationen verwendet. Die Ergebnisse deuten darauf hin, dass die Netzwerkstruktur als Ganzes relevante biologische Informationen erfasst.

Abschließend beschreibe ich die öffentlich zugängliche Benutzeroberfläche, die ich implementiert habe, um einen effizienten Zugriff auf die AD Atlas-Daten zu ermöglichen. Diese netzwerkbasierte Plattform ermöglicht es Forschern eigene Hypothesen zu formulieren und klinisch relevante Ergebnisse abzurufen, die in Folgeanalysen oder Experimenten validiert werden können. Anhand mehrerer Beispiele demonstriere ich den Nutzen und die Relevanz dieser Ressource für die Alzheimer-Forschung. Anhand von zwei bekannten Alzheimer-Risikogenen zeige ich beispielsweise, wie der AD Atlas verwendet werden kann, um Kandidaten für die Neupositionierung von Medikamenten zu identifizieren. Mit diesem Ansatz werden zwei Wirkstoffe, Candesartan und Levetiracetam, identifiziert, die derzeit in klinischen Studien getestet werden, sowie mehrere andere Kandidaten, die ebenso vielversprechend sein könnten und zusätzliche Erkenntnisse liefern können.

Der vorgestellte AD Atlas ist eine wichtige neue Multi-omics-Ressource, die eine globale molekulare Ansicht auf Alzheimer bietet und eine Vielzahl an Analysewerkzeugen bietet, um auf molekulare Daten basierende Hypothesen zu kontextualisieren. Die in dieser Arbeit vorgestellten Konzepte und Erweiterungen sind vielseitig anwendbar und bieten einen globalen Rahmen für die integrative Analyse auch anderer komplexer Krankheiten.

Scientific contributions

Key publications

This dissertation is based on the following publications that have either been published in peer-reviewed journals or are currently being prepared for submission.

- M. A. Wörheide, J. Krumsiek, S. Nataf, K. Nho, A. K. Greenwood, J. C. Wiley, T. Wu, K. Huynh, P. Weinisch, W. Römisch-Margl, N. Lehner, The AMP-AD Consortium, The Alzheimer's Disease Neuroimaging Initiative, The Alzheimer's Disease Metabolomics Consortium, J. Baumbach, P. J. Meikle, A. J. Saykin, P. Murali Doraiswamy, C. van Duijn, K. Suhre, R. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"An online molecular atlas of Alzheimer's disease"**, *submission in preparation*. Preprint available on medRxiv, doi: 10.1101/2021.09.14.21263565.
- M. A. Wörheide, J. Krumsiek, S. Nataf, K. Nho, A. K. Greenwood, J. C. Wiley, T. Wu, K. Huynh, P. Weinisch, W. Römisch-Margl, N. Lehner, The AMP-AD Consortium, The Alzheimer's Disease Neuroimaging Initiative, The Alzheimer's Disease Metabolomics Consortium, J. Baumbach, P. J. Meikle, A. J. Saykin, P. Murali Doraiswamy, C. van Duijn, K. Suhre, R. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"Utilizing multi-omics context networks to explore molecular hypotheses in Alzheimer's disease "**, *submission in preparation*. Preprint available on medRxiv, doi: 10.1101/2021.09.14.21263565.
- M. A. Wörheide, J. Krumsiek, G. Kastenmüller, and M. Arnold, **"Multi-omics integration in biomedical research – a metabolomics-centric review"**, *Anal. Chim. Acta*, vol. 1141, pp. 144–162, Jan. 2021.

Additional contributions

The following list provides additional scientific contributions through peer-reviewed publications or talks or poster presentations at international conferences sorted in descending order by date.

Publications

- P. Surendran, I. D. Stewart, V. P. W. Au Yeung, M. Pietzner, J. Raffler, M. A. Wörheide, C. Li, R. F. Smith, L. B. L. Wittemans, L. Bomba, C. Menni, J. Zierer, N. Rossi, P. A. Sheridan, N. A. Watkins, M. Mangino, P. G. Hysi, E. Di Angelantonio, M. Falchi, T. D. Spector, N. Soranzo, G. A. Michelotti, W. Arlt, L. A. Lotta, S. Denaxas, H. Hemingway, E. R. Gamazon, J. M. M. Howson, A. M. Wood, J. Danesh, N. J. Wareham, G. Kastenmüller, E. B. Fauman, K. Suhre, A. S. Butterworth and C. Langenberg, **"Rare and common genetic determinants of metabolic individuality and their effects on human health"**, *Nat. Med.*, vol. 28, pp. 2321-2332, Nov. 2022.

- R. Batra, M. Arnold, M. A. Wörheide, M. Allen, X. Wang, C. Blach, A. I. Levey, N. T. Seyfried, N. Ertekin-Taner, D. A. Bennett, G. Kastenmüller, R. F. Kaddurah-Daouk, J. Krumsiek, and Alzheimer's Disease Metabolomics Consortium (ADMC), **"The landscape of metabolic brain alterations in Alzheimer's disease"**, *Alzheimers. Dement.*, pp. 1-19, doi: 10.1002/alz.12714, July 2022.

- M. Pietzner, E. Wheeler, J. Carrasco-Zanini, A. Cortes, M. Koprulu, M. A. Wörheide, E. Oerton, J. Cook, I. D. Stewart, N. D. Kerrison, J. Luan, J. Raffler, M. Arnold, W. Arlt, S. O'Rahilly, G. Kastenmüller, E. R. Gamazon, A. D. Hingorani, R. A. Scott, N. J. Wareham, and C. Langenberg, **"Mapping the proteo-genomic convergence of human diseases"**, *Science*, vol. 374, p. eabj1541, Nov. 2021.

Talks

- **"From associations to insights? - Building an integrated molecular atlas of Alzheimer's disease"**, Talk at the 10th *Grainau Workshop Genetic Epidemiology*, May 2022.

- **"Application of the AD Atlas: integrating the universe of molecular Alzheimer's data with pseudo-temporal models"**, Talk at the *National Institute on Aging (NIA) – Annual Tri-Consortia Meeting*, Dec 2021.

- **"A proof of concept study towards multi-omics-based computational drug repositioning in Alzheimer's disease"**, Talk at the *Alzheimer's Association International Conference*, July 2021.
- **"A network-based, multi-omics Atlas for target identification and prioritization in Alzheimer's disease"**, Talk at the *Alzheimer's Association International Conference*, July 2020.

Poster presentations

- M. A. Wörheide, R. Batra, J. Krumsiek, R. F. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"Multi-omics characterization of brain-based pseudotime estimates for Alzheimer's disease progression"**, Poster presentation at *Alzheimer's Association International Conference*, July 2022.
  Commented on/covered by Alzforum: "Pseudotime Simulates Disease Progression from Case-Control RNA-Seq", alzforum.org/news/conference-coverage/pseudotime-simulates-disease-progression-case-control-rna-seq.
- M. A. Wörheide, J. Krumsiek, G. Kastenmüller, and M. Arnold, **"A network-based approach for the identification of multi-omics modules associated with complex human diseases"**, Poster presentation at *EMBO | EMBL Symposium: Multiomics to Mechanisms - Challenges in Data Integration*, Sep. 2019. Received a best poster award.

Contents

Overview and objectives 1

In recent years, technological and computational advances have propelled molecular biology into a high-throughput and big-data era. It has become possible to simultaneously measure hundreds to thousands of molecules in a biological sample. This has provided researchers with highly valuable, so-called 'omics', datasets, including genetic sequence data (genomics, transcriptomics), protein abundances (proteomics), and metabolite levels (metabolomics). In recent years the integrative analysis of different data modalities has become increasingly important as an addition to single omics studies. In so-called multi-omics analyses, researchers combine different omics datasets (e.g., genomics, transcriptomics, and metabolomics) to gain a more comprehensive picture of the molecular mechanisms of biological systems and pinpoint potential causative changes that lead to disease [1]. Especially for complex, heterogeneous, and untreatable diseases such as AD, integrative analyses are desperately needed to further understand disease mechanisms and ultimately identify novel therapeutic targets. However, large-scale multi-omics studies still face critical challenges, such as inadequate sample sizes and the lack of universal protocols and standardized analysis pipelines [2].

International and interdisciplinary research efforts have made significant investments to thoroughly characterize large cohorts, provide high-quality quantitative omics datasets, and share them with the broader research community [3, 4, 5]. Nonetheless, integrative analyses are complex, partly due to the high degree of data heterogeneity, which includes different data formats (CSV, Excel, JSON), and types (quantitative vs. qualitative). In addition, not all omics modalities are necessarily available for the same set of samples, for example, due to budget restrictions or sample availability. Consequently, each data set requires different processing, and careful consideration must be given to determine how to interconnect the existing datasets best. Another challenge is the high complexity of relationships between and across different omics entities (e.g., genes, transcripts, and metabolites) and finding comprehensible representations for these vast amounts of data [6]. To this end, networks provide an intuitive and mathematically-well defined framework to store and interconnect diverse biological domains [7, 8]. They enable the application of established graph algorithms, such as module identification [9], and can be simplified to

provide more interpretable views by operations, such as edge contraction, i.e., joining nodes after the removal of an edge. However, building such an integrative network view from heterogeneous data requires extensive domain knowledge. Furthermore, significant effort is needed to develop appropriate data structures efficient enough for downstream analysis and flexible enough to allow the inclusion of new data types [8].

Lack of (bio)informatic expertise, missing domain knowledge, and insufficient computational resources present a bottleneck in translating heterogeneous scientific data into insights and, ultimately, new therapeutic approaches. Therefore, there is an urgent need for research frameworks and resources that facilitate the integration of heterogeneous omics data and that enable researchers - with or without a computational background - to leverage existing data and generate new insights through integrative analyses.

1.1 Scientific aims

This thesis aims to provide an overview of current multi-omics integration strategies and to develop a novel network-based multi-omics framework that enables the large-scale integration and analysis of heterogeneous (AD-related) data. The following objectives summarize the aims of this work:

1. Provide an overview and methodologically categorize current multi-omics data integration strategies and workflows that can combine metabolomics data with other omics, emphasizing their application potential, strengths, and drawbacks.
2. Develop a network-based framework that allows the large-scale integration and analysis of heterogeneous molecular data collected across cohorts for AD research.
3. Implement a flexible user interface that supports the full exploration, visualization, and analysis of multi-omics data through the generation of context-specific molecular networks.
4. Demonstrate the usability and relevance of the developed resource for AD research.

1.2 Outline

The content of this work is visually outlined in Figure 1.1 and is structured as follows:

Chapter 2 provides an in-depth introduction to and discussion of current multi-omics integration methods, particularly focusing on the integration of metabolomics data. I introduce integration methods that enable researchers to analyze large-scale, heterogeneous biological datasets and identify their respective application potential and weaknesses. Furthermore, I explore simultaneous and step-wise integration strategies and categorize existing methods into knowledge-based, data-driven, and composite network approaches.

Chapter 3 introduces the neurodegenerative disorder AD, providing an overview of disease hallmarks, current research challenges, and existing multi-omics resources.

Chapter 4 provides an overview of key AD cohorts, datasets, and methods that I used in the development of the AD Atlas – an integrative, multi-omics AD resource presented in Chapter 5. In addition, this Chapter introduces analytical concepts for the downstream analysis of multi-omics networks that were used throughout this work and provides details on the technical framework of the resource and the accompanying user interface.

Chapter 5 presents the AD resource that I built using the developed integration framework. The AD Atlas is a network-based data integration resource for studying AD, its biomarkers, and related (endo-)phenotypes in a multi-omics context. I used an extended quantitative trait loci (QTL)-based integration method paired with a composite network approach to combine data from over 25 large-scale studies. To enable access to this resource, I built a user-friendly, network-based interface where users can explore entities (metabolites, genes, AD-phenotypes) of interest in a multi-omics context through provided tools. Finally, I demonstrate the usability and relevance of the AD Atlas for AD research through multiple showcases covering both hypothesis-driven and exploratory research questions.

Chapter 6 discusses current multi-omics integration strategies and the value of the AD Atlas as a multi-omics research resource. Furthermore, I highlight the potential of this framework for AD research and other complex diseases. Lastly, I critically reflect upon current limitations and future perspectives for the AD Atlas.

Chapter 7 concludes this work with a summary, highlighting my scientific contributions to AD multi-omics research.

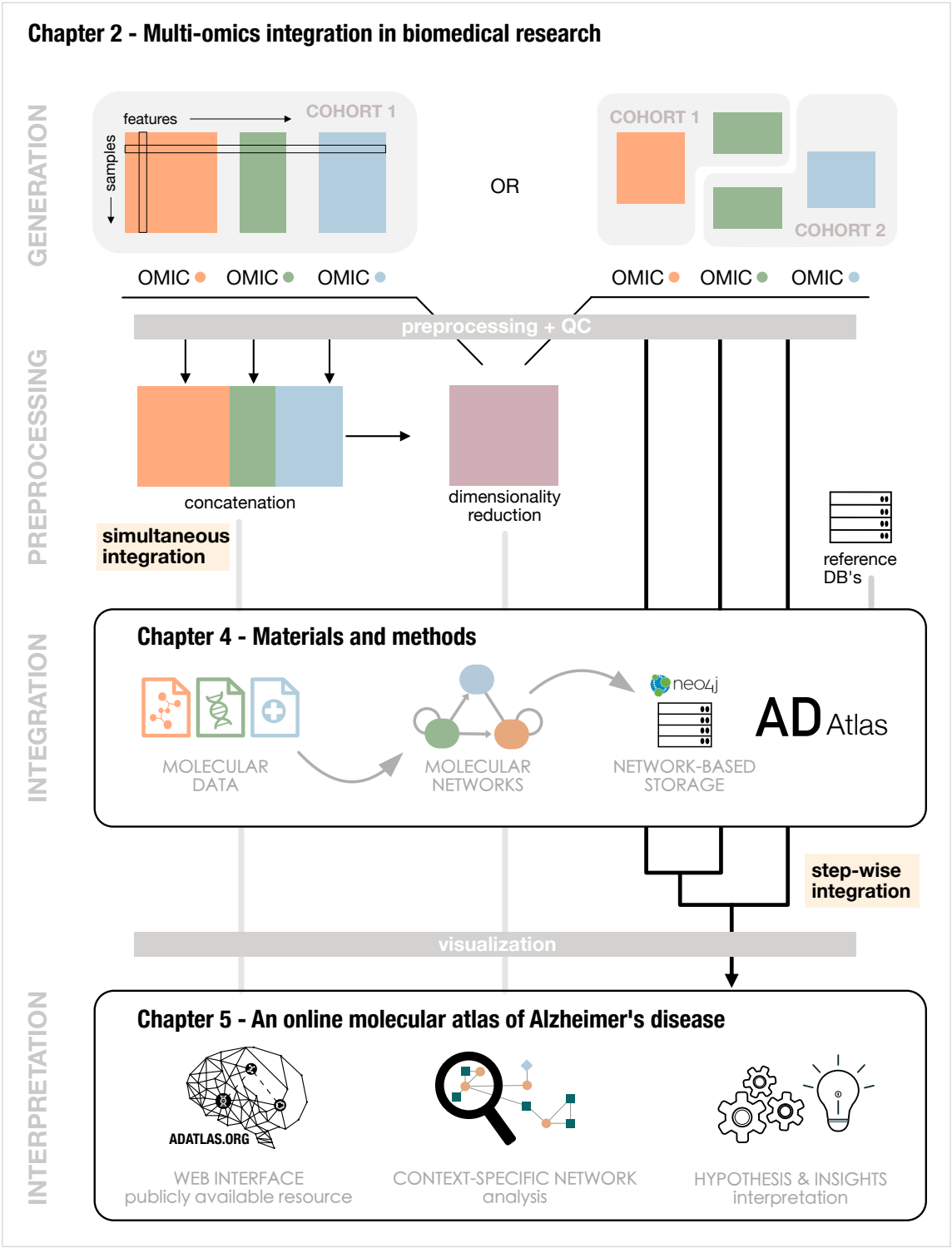


Figure 1.1: General thesis overview. First, I briefly motivate this work and provide my scientific aims (**Chapter 1**). Next, I introduce and discuss different omics integration strategies, highlighting their strengths and weaknesses (**Chapter 2**). **Chapter 3** gives an introduction to AD and in **Chapter 4** I detail key AD cohorts, datasets and methods used throughout this work. I introduce the network-based, multi-omics integration platform – the AD Atlas – in **Chapter 5** and provide a discussion with future outlooks in **Chapter 6**. Lastly, I conclude this work in **Chapter 7**. Figure adapted from Wörheide et al. [10].

Multi-omics integration in biomedical research 2

A similar version of this Chapter has been published in [10]:

M. A. Wörheide, J. Krumsiek, G. Kastenmüller, and M. Arnold, **"Multi-omics integration in biomedical research – a metabolomics-centric review"**, *Anal. Chim. Acta*, vol. 1141, pp. 144–162, Jan. 2021.

Contributions of the lead author were as follows:

- conceptualization of the review (in consultation with co-authors)
- literature review
- drafting the manuscript
- design and implementation of all figures

Critical revision of the article and final approval was given by all.

Advances in high-throughput technologies have enabled the generation of vast amounts of data on multiple layers of a biological system, including deoxyribonucleic acid (DNA) sequence data (genomics), ribonucleic acid (RNA) expression levels (transcriptomics), epigenetic alterations (epigenomics), protein abundances (proteomics), metabolite levels (metabolomics) and more. Considering each of these biological layers separately, numerous omics studies identified genes, proteins, and metabolites that associate with specific diseases or phenotypes of interest. For example, high levels of branched-chain amino acids (BCAAs) and their degradation products have been found as hallmarks of type 2 diabetes [11]; in contrast, Alzheimer's disease (AD) associates with low levels of these metabolites [12]. While the identified entities can serve as valuable biomarkers and provide insights into pathways involved in pathomechanisms, single omics studies do not take into account the complex interplay of various biological layers. However, disturbances of cross-omics interactions might play important roles in the development and clinical presentation of a disease [1, 13]. Therefore, combining omics data from multiple biological domains

(e.g. levels of transcripts, proteins, or metabolites) in multi-omics studies is a promising approach toward a more detailed molecular understanding of health and disease, as well as the chain of cause and effect, which is an essential requirement for guiding novel therapies [14]. For example, results from an integrated analysis of large genetic and metabolomic datasets by Lotta et al. [11] using a Mendelian Randomization approach, were consistent with a causal role of BCAA metabolism in type 2 diabetes and suggested the *PPM1K* gene (genetic variants therein being specifically associated with levels of BCAAs in the blood) as a potential drug target. *PPM1K* encodes the mitochondrial phosphatase that activates the branched-chain alpha-ketoacid dehydrogenase (BCKD) complex, the rate-limiting enzyme in BCAA catabolism, and was only up-regulated in muscles of healthy subjects but not in patients with type 2 diabetes in a validation experiment. Although the availability of multi-omics data does not always allow for direct conclusions on causality, the combination of multiple layers of evidence in a multi-omics study has been demonstrated to provide more reliable results and mitigate the risk of false positive findings [15, 16]. Beyond the value of multi-omics approaches for the investigation of particular diseases, large-scale multi-omics studies enable the systematic investigation of inter- (e.g. enzymatic conversion of metabolites) and intra-omics (e.g. protein-protein interactions) relationships independent of a specific phenotype.

In multi-omics studies, metabolomics occupies a unique position and has received increasing attention in integrative analysis [17]. Metabolites are the downstream output of biological processes, carrying imprints of genomic, epigenomic, and environmental effects. They are often referred to as 'the link between genotype and phenotype' [18] and have been implicated in numerous diseases, such as AD [19], type 2 diabetes [20], and various types of cancer [21]. Furthermore, they carry integrated biological and medical signals in easily accessible biofluids (e.g., blood, urine), making them attractive biomarker candidates [22]. Large-scale epidemiological studies have demonstrated the value of integrating metabolomics with other omics layers, such as genomics [23,24,25,26], transcriptomics [27] and epigenetics [28], providing insight into metabolic individuality and links to disease mechanisms [29, 30]. For example, up to 62 percent of variation in metabolite concentration levels in two population-based cohorts could be explained by common genetic variants [25]. Furthermore, it has been shown that DNA methylation affects metabolism [31]. This effect is partly driven by genetic variation, but further depends on environmental and lifestyle factors, enabling an adaptive response to regular (e.g., food intake) [32, 33] and specific (e.g., disease) [34] challenges. Changes in the metabolome can, in turn, modulate the activity of genes and proteins, creating complex feedback mechanisms and interrelationships between omics layers [35]. Therefore, the integration of metabolomics with other omics layers provides exciting opportunities for the study of disease mechanisms and the identification of novel therapeutic targets.

Box 2.1: Important terms and concepts.*

Integration method - A specific method/framework that performs data integration.

Integration strategy – Summary term for multiple data integration methods that follow the same principle.

Knowledge-based integration – Relationships between biological entities across and within omics are established using knowledge bases (extrinsic information).

Data-driven integration – Relationships between biological entities across and within omics are statistically inferred from multi-omics datasets (intrinsic information).

Simultaneous integration – Integration strategies that take into account all available data by merging the data and performing a single method on the concatenated matrix.

Single-block methods – Multivariate methods that perform simultaneous integration and do not take into account heterogeneities between the different omics datasets.

Multi-block methods – Multivariate methods that perform simultaneous integration and can take into account the block structure of multi-omics data by modeling each block separately.

Step-wise integration – Integration strategies that analyze omics datasets separately and integrate the results or models in a subsequent step.

Biological entity – Refers to a measured biological molecule such as protein, metabolite, or lipid but also includes single nucleotide polymorphism (SNP) and epigenetic alterations.

*Box reprinted from [10].

To enable the analysis of heterogeneous datasets in multi-omics studies, a plethora of data reduction, manipulation, and integration techniques have been developed. Previous review articles have provided comprehensive method summaries for specific integration strategies such as network inference and analysis [36,37] or machine learning techniques [38,39,40,41], and have discussed important aspects of metabolite-centered studies [2, 42, 43]. However, most work concentrates on the integration of two different data types with respect to a specific phenotype of interest. Therefore this Chapter will provide an overview of a typical multi-omics workflow, focusing on integration methods that have the potential to combine metabolomics data with more than two omics and highlighting their application in recent multi-omics studies. We will distinguish between integration efforts that build prediction models [44, 45, 46, 47] or identify diagnostic and prognostic biomarker candidates [47, 48] for a specific disease phenotype or trait of interest, and global integration efforts that are initially not focused on a specific outcome. The latter approaches aim at the systematic

integration of multiple omics datasets to provide a basis for generating testable hypotheses and gaining mechanistic insights into the pathophysiology of multiple complex diseases in post-integration analyses [49, 50, 51].

The choice of an appropriate integration strategy is not straightforward and heavily depends on the available data and study objective. Data dimensionality, heterogeneity, and lack of universal protocols additionally complicate this task. Generally, two major integration paradigms (Figure 2.1) have been described in the literature [2, 36, 42, 52, 53] and will be referenced throughout this Chapter; (1) simultaneous and (2) step-wise integration. *Simultaneous integration* strategies use all available omics data at the same time and perform analysis in a single modeling step. Thereby, complementary information encoded in each omics layer, as well as correlations between the layers, are taken into account. Methods of this category require that the data was derived from the same biological samples or individuals, which poses still a major limitation regarding the availability of such data due to funding or technical restrictions. *Step-wise integration* strategies, on the other hand, analyze omics datasets in isolation or in specific combinations and integrate the results in a subsequent step. This facilitates the integration of data and statistical results from different sources (e.g., different studies or knowledge bases), allowing the large-scale analysis of heterogeneous data in the absence of omics measurements for the same samples.

This Chapter will introduce central aspects of a typical multi-omics data integration workflow (Figure 2.1) and is structured as follows: (i) *Data scenarios.* Study design, sample preparation, and subsequent data acquisition through high-throughput analytical platforms can lead to different data scenarios. (ii) *Dimensionality reduction.* After appropriate preprocessing of raw data collected on different omics layers, dimensionality reduction is often applied to reduce the number of variables (measured biological entities). (iii) *Data integration.* Data from different omics layers are analyzed and integrated using a method that is appropriate for the input data and research question of interest. (iv) *Data interpretation.* Post-integration inspection and further analysis of the integration results (e.g., statistical model or network) enable meaningful biological insights. A short outlook on future directions for multi-omics research concludes this Chapter.

2.1 Data scenarios

Integrative multi-omics analyses combine several omics measurements, optionally along with additional phenotypes of interest, that are represented by either continuous (e.g., protein levels or metabolite concentrations) or categorical variables (e.g., gender or disease status). Naturally, each dataset comes in a separate data matrix where rows represent individual samples, and columns hold measurements of demographic, clinical, or biological entities (Figure 2.1). However, depending on the study objective and access to relevant data, there are three different data scenarios: (1) the different datasets are available for the

same samples/individuals; (2) the datasets are available for only a partially overlapping set of samples/individuals; (3) omics data is distributed across mostly disjoint sets of samples.

In the first scenario, samples from a study are simultaneously subjected to the same multi-omics screening processes, or additional omics technologies are applied to initially collected samples in retrospect. Data from such studies will result in data matrices where the rows in every data matrix correspond to the same samples/individuals and columns hold measurements for each respective omics technology (e.g., metabolomics, transcriptomics, proteomics). This is the optimal scenario, as it allows the application of any integration strategy, including simultaneous data integration that requires data matrices with matched samples [54].

However, complete multi-omics profiles are often not available or feasible to get for all samples/study participants. The reasons for this are manifold and include funding limitations, incompatibility of collected samples for certain omics analyses, or depletion of samples preventing the application of novel technologies [2, 54]. For example, although urine samples have proven very informative in metabolomics studies, they contain limited amounts of proteins and RNA, limiting their use in large-scale proteomics or transcriptomics studies [2]. Furthermore, in long-term studies or studies with rollover participants, both omics and phenotypic screenings applied at baseline may be adapted due to technological advances, falling costs for sample analysis, or evolving study objectives. For instance, the Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal, multicenter study launched in 2004 to study biomarkers for early detection of AD [55]. While large-scale metabolomics and lipidomics profiling is available for the study phases ADNI-1 and -GO/2, up to now (biosamples are still available) proteomics profiling was only applied to a subset of ADNI-1 participants and gene expression profiling is only available for ADNI-GO/2. This leads to differing availabilities of omics profiles for participants across study phases.

Data resulting from such a study will only have partially overlapping samples for multi-omics integration [54]. If the overlap of samples between data types is large enough for a sufficiently powered study, the removal of samples without full omics profiles can still enable simultaneous integration. However, the application of such a list-wise deletion of individuals is prone to substantial loss of information [39, 54]. In the worst case, this can introduce estimation bias by resulting in a sample set that is unrepresentative of the initial study population [56]. Nevertheless, simultaneous data integration strategies are emerging that can handle a moderate amount of samples with missing omics profiles (see Section 2.3.2).

Due to the restrictions mentioned above, many multi-omics analyses use datasets that have not been collected from the same samples and originate from different sources. A special case of this scenario occurs if the sample sets for each data type were acquired in the same study but have minimal overlap. By integrating such omics measurements, data

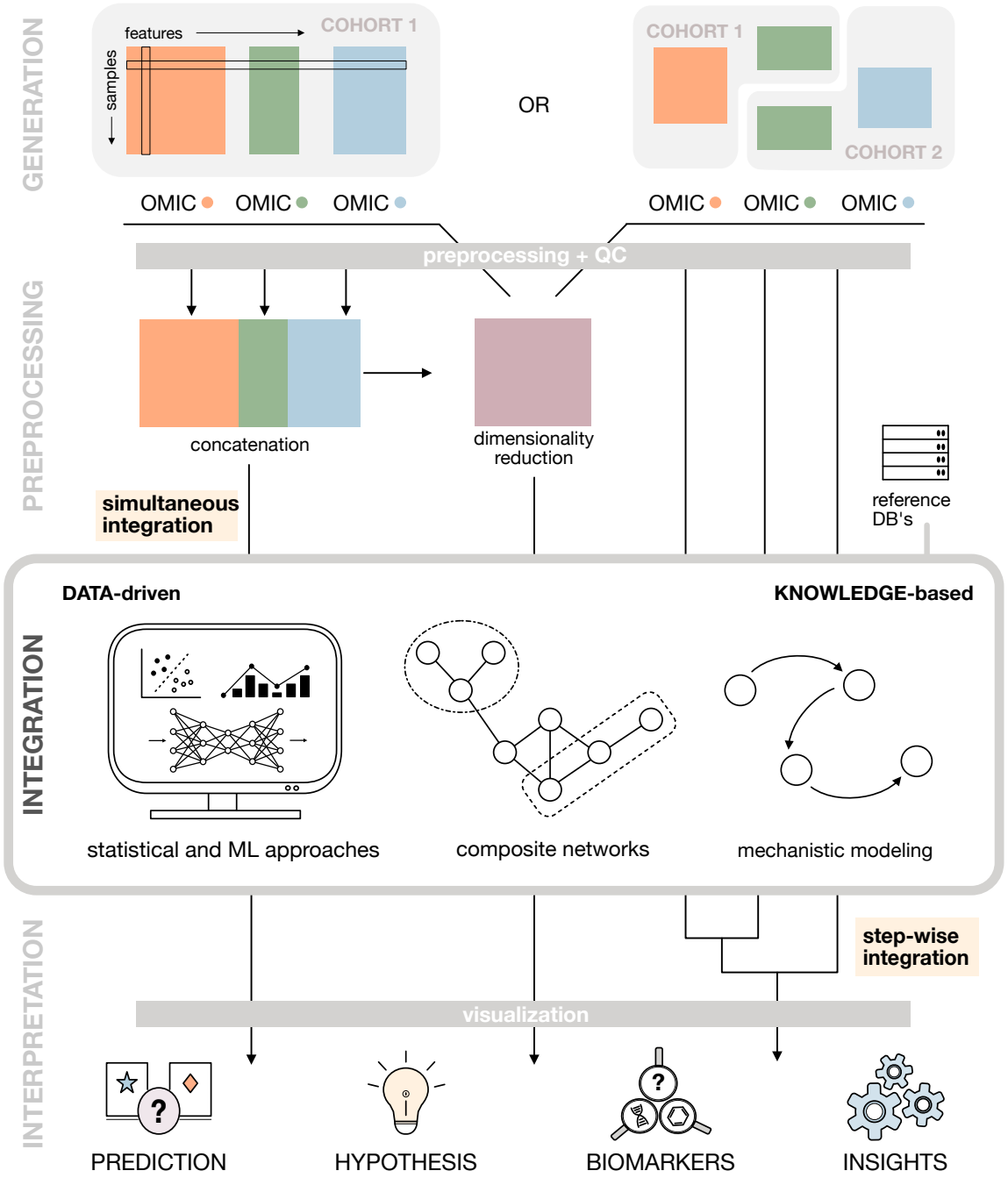


Figure 2.1: Multi-omics workflow. A typical multi-omics analysis can generally be broken down into four steps. **(i) Data generation.** Study design, sample preparation, and subsequent data acquisition through high-throughput analytical platforms lead to different data scenarios. **(ii) Data preprocessing and dimensionality reduction.** Raw data collected on different omics layers is preprocessed appropriately and dimensionality reduction can be applied to reduce the number of variables (measured biological entities). **(iii) Data integration.** Data from different omics layers are analyzed and integrated using data-driven, knowledge-based or hybrid integration approaches. The choice of method depends on the input data and research question of interest. **(iv) Data interpretation.** Post-integration visualization and analysis of the integration results (e.g., statistical model or network) can identify novel biomarker candidates, generate testable hypotheses or reveal meaningful biological relationships. Figure and caption reprinted from Wörheide et al. [10].

matrices consequently have mostly unmatched samples and variables as a starting point. For this data scenario, several step-wise integration strategies (discussed in Sections 2.3.1 and 2.3.2) have been developed that enable both multi-omics analyses in disjoint sample sets and inclusion of preexisting biological data. However, it is important to keep in mind that these types of analyses add another layer of data heterogeneity due to differing sample sizes, study protocols, and study demographics (e.g., age, sex, or ethnicity).

In summary, multi-omics datasets available for the same samples/individuals introduce less unwanted data heterogeneity and enable the application of any integration method. For datasets with only partially overlapping or completely disjoint sets of samples/individuals, the number of applicable integration methods is a bit more limited, but those that are available allow for almost infinite inclusion of data, enabling studies to yield maximal power.

2.2 Dimensionality reduction

Appropriate preprocessing of the raw data is a key prerequisite for any type of analysis, as technical artifacts and skewed data distributions can distort biological signals [57]. This process typically includes the removal of batch effects, normalization, and imputation of missing values for each data type separately before integration [6]. The importance of study design and temporal ordering of sample collection [2, 6, 42, 58], as well as guidelines for appropriate data preprocessing [6, 39], have been discussed in previous reviews and are beyond the scope of this work. In the following, we will assume that the data that is subjected to integrative analyses was appropriately preprocessed and is of high quality.

The curse of dimensionality [59] is a central challenge in single-omics studies and even further aggravated in multi-omics studies, where the number of variables is substantially higher. With increasing dimensions (number of variables), distance measures become meaningless, which is challenging for operations in this high-dimensional space, such as clustering [60, 61]. Furthermore, samples are typically significantly outnumbered by measured variables, posing a challenge for most statistical learning methods. This can lead to an underdetermined mathematical system and increases the risk of overfitting classifiers or predictors [36]. Dimensionality reduction (DR) is a way to reduce the complexity of a dataset while increasing prediction stability, boosting the statistical power of downstream analyses, and reducing the multiple testing burden. DR is performed by either extracting relevant variables (feature selection) or projecting data onto a lower-dimensional space (feature extraction) [39].

Feature selection often involves prior knowledge or a biological hypothesis that is used to reduce the number of considered variables. Popular approaches are, for example, to limit the analyses to genes, proteins, and metabolites involved in certain pathways of interest, or to investigate entities that have been previously associated with a specific trait

under study [49]. Such hypothesis-driven DR strategies can significantly boost statistical power but are naturally prone to bias towards biological entities that have been annotated through previous studies. Another knowledge-based approach is to construct new variables that are biologically meaningful, i.e., representative of functional groups such as pathways. For example, metabolites can be analyzed at the pathway level by aggregating levels of all molecules assigned to a specific pathway (e.g., by using the average z-score of concentrations [62] or first principal component from a Principal Component Analysis (PCA) [62, 63, 64]) to produce new pathway-based variables [65].

Feature extraction, on the other hand, is typically achieved by data-driven DR techniques such as PCA [39, 66]. PCA is classically applied to each omics dataset separately and transforms single-omics variables into a lower-dimensional subspace that maximizes the retained variance within the data by finding orthogonal linear combinations of the original variables. Therefore, PCA enables the use of a reduced set of features with minimal loss of information. Related approaches include clustering techniques (e.g., K-means [67] or hierarchical clustering [68]) followed by the replacement of groups of similar variables by a cluster centroid [69]. Here, one popular approach is to cluster correlating biological entities, such as metabolites, proteins, or transcripts, by using weighted gene co-expression network analysis (WGCNA) [70] on each dataset [71, 72]. The identified clusters are then summarized by the first principal component from a PCA ('eigengene' or 'eigenmetabolite') on the abundance matrix of each respective cluster that is then used in downstream analyses (e.g. association with a specific phenotype, integration with other omics layers) with a reduced set of features [73]. A limitation of such data-driven approaches is that the interpretation of the derived associations or correlations requires the extracted features to be mapped back onto the original variables.

In summary, DR provides a way to limit the potential for overfitting and significantly reduces the multiple testing burden. Additionally, knowledge-based DR can increase the downstream interpretability of analysis outcomes.

2.3 Data integration

The growing interest in integrative analysis of multi-omics datasets has led to the emergence of various integration frameworks. In the following, the major concepts are introduced and categorized into approaches that take into account external information (knowledge-based approaches) and approaches that primarily rely on intrinsic information (data-driven approaches) to infer dependencies across omics. Finally, we will turn to hybrid approaches (composite networks) that combine knowledge-based and data-driven integration.

2.3.1 Knowledge-based approaches

Knowledge-based integration strategies use external information from databases or scientific literature to establish relationships between biological entities. Results from previous analyses are either annotated using prior knowledge (e.g., using common functional terms) or mapped onto a reference network that connects different omics layers based on established knowledge. For example, metabolic networks assembled based on biochemical knowledge, enable the connection of enzymes and metabolites through reactions. By mapping results from single-omics analyses onto such a network, findings can be integrated and interpreted in a multi-omics context, enabling the identification of pathways that are dysregulated at the gene, protein, and metabolite level [74]. Furthermore, multi-omics measurements can be integrated into preexisting biological models to make them condition-specific (e.g., deletion of inactive reactions) [75].

Prior knowledge that is used for this type of omics integration includes, but is not limited to, information on functional relationships (e.g., pathways or biological reactions), pharmacogenomic associations, and genome annotations. Depending on the source, this information is either based on experimental data [76], collected from scientific literature (manually or by using automated text-mining techniques) [77], or derived from computational prediction approaches [78]. As knowledge bases typically combine information from multiple sources, they can have varying levels of evidence. For example, STRING [77], a popular protein-protein interaction (PPI) database, indicates the confidence of functional interactions between proteins by assigning scores that are based on the quality and type of supporting evidence coming from targeted experiments, co-expression analysis, genomic context predictions, or text-mining [79].

While many resources are specific to one omics type, such as STRING or the LIPID MAPS Structure Database (LMSD) [80] for lipid annotations, a number of databases have emerged that cover multiple biological domains (see Table 2.1). The Kyoto Encyclopedia of Genes and Kyoto Encyclopedia of Genes and Genomes (KEGG) [81,82,83] database, for instance, was released in 1995 as one of the first computational resources that linked the genome with higher-order functional information. In KEGG, manually compiled pathway maps enable researchers to view genes and proteins in the context of metabolic networks and pathways, such as sphingolipid metabolism or NF-kappa B signaling. Nearly a decade later, additional curated and pathway-centered resources started emerging, such as Reactome [84,85] and Recon [86, 87, 88]. Reactome is a resource that is primarily focused on human biological processes and is built around reactions. Reactions are defined as an event that transforms an input to output (both being biological entities such as proteins, lipids, or nucleotides) and are further grouped into pathways depending on their (temporal) relationships [84]. Taking this concept a step further, Recon3D [86,87,88] provides a genome-scale metabolic reconstruction that can be used for computational modeling (see Section 2.3.1 on constraint-based metabolic modeling). It also includes three-dimensional (3D) structural

data on metabolites and proteins and represents the most comprehensive human metabolic network model to date [88].

In order to utilize these resources for knowledge-based integration, platform-specific identifiers (IDs) of measured biological entities need to be mapped to the namespace of the respective target database. This task is challenging, as most resources have developed their own internal ID schemes and hierarchies, leading to a plethora of IDs across databases that refer to the same entity. Efforts have been made to enable cross-linking between ID schemes [88] and mapping tools are available online or through R packages, such as biomaRt [89] for genes or MetaboAnalystR [90, 91] for metabolites. However, name ambiguities, ID multiplicity, and the use of synonyms complicate this task [92] and can lead to significant loss of information if not handled carefully. This is especially challenging for metabolites and lipids due to differences in resolution between platforms and technologies [93]. For example, lipid sidechain composition and configuration are important determinants of the function of phosphatidylcholiness (PCs). However, many lipidomics techniques cannot distinguish between isobaric species sharing the same nominal mass [94] and annotate PCs at the lipid species level assuming even-numbered fatty acids, as they are more frequent, i.e., PC (731) with m/z 731 will most likely be labeled PC 32:1 and not PC O-33:1, although both are plausible [93].

Knowledge bases are under constant pressure to adapt to technological advances and incorporate novel research findings (e.g., the discovery of various types of regulatory RNA species) to accurately reflect the current state of science, which can lead to further discrepancies. For example, despite the fact that some platforms offer fatty acid side-chain resolving techniques, lipids are often not yet annotated at this level of detail [15] and this information will be lost when matching measured compounds to the namespace of a resource (e.g., PC 16:0_16:1 would simply be mapped to the KEGG identifier C00157 for phosphatidylcholine).

Nevertheless, when correctly employed knowledge bases provide a wealth of valuable information that can be exploited in multi-omics integration.

Set-based enrichment

Set-based enrichment is a commonly used, step-wise results integration strategy. It tests whether certain functional annotations are enriched in a list of interesting (e.g., differentially expressed or abundant) biological entities, which have been identified in preceding omics analysis. Biological entities are assigned to sets (also referred to as annotation terms) using information from knowledge bases to examine whether they are known to participate in the same biological pathways, are significantly changed in a specific disease, or are co-localized (e.g., in the same organelles, tissues or organs) [109]. For example, the annotation term 'sphingolipid metabolism' in Reactome [84, 85] includes metabolites, such as sphin-

Table 2.1: A selection of network-based multi-omics knowledge bases, visualization tools, and online resources. Table and caption reprinted from Wörheide et al. [10].

	Network visualization	Analysis tools	Project omics data onto network	Biological entities	Implementation	Reference
BioCyc	x	Enrichment analysis Flux analysis	x	genes proteins metabolites	online	[95]
KEGG	x	-	x	genes enzymes metabolites	online KEGGscape^{+} CytoKegg^{+}	[82]
Reactome	x	Enrichment analysis ID mapping	x	proteins metabolites diseases	online Reactome FIViz^{+}	[84]
Recon3D	x		x	genes metabolites	online	[88]
PathwayCommons	x	Enrichment analysis	-	proteins metabolites drugs	online R CyPath2^{+}	[96]
WikiPathways	x	-	-	genes proteins metabolites	online WikiPathways^{+}	[97, 98]
NDEx	x	Neighborhood search	-	various**	online CyNDEx-2^{+}	[99, 100, 101]
PaintOmics3	x	Clustering Correlation analysis Enrichment analysis ID mapping	x	genes proteins metabolites	online	[74]
MetaboAnalyst	x	Enrichment analysis ID mapping Shortest path analysis	x	genes metabolites	online R	[90]
OmicsNet	x	Clustering Enrichment analysis Shortest path analysis	x	genes proteins TFs miRNAs metabolites	online	[102]
MetExplore	x	Enrichment analysis Flux analysis ID mapping Shortest path analysis	x	genes enzymes metabolites	online	[103]
ConsensusPathDB	x	Clustering Enrichment analysis Shortest path analysis	x	genes proteins metabolites	online	[104]
PathMe Viewer	x	Shortest path analysis	-	genes proteins metabolites	online	[105]
MetScape	x	Correlation analysis Enrichment analysis	x	genes enzymes metabolites	MetScape^{+}	[106, 107]

**no restrictions $^{+}$Cytoscape Application [108]

gosine 1-phosphate and sphingosine, and genes such as *SGPP1* (sphingosine-1-phosphate phosphatase 1) and *SPHK1* (Sphingosine Kinase 1). Here, we focus on the most widely used approaches: overrepresentation analysis and functional set enrichment analysis.

Over-representation analysis (ORA) aims at the identification of annotation terms that are overrepresented, i.e. terms that are more frequently assigned to the entities in the input list of interest than expected by chance [109]. This can be statistically tested by using a hypergeometric test such as one-sided Fisher's exact test with subsequent correction for multiple testing [109]. In order to yield meaningful results, a valid definition of the background, i.e., the set of entities that were measured in the analysis and assigned to each annotation term, is a key requirement [58] in order to correct for bias

that arises due to unequal annotation coverage of different entities. This is a prominent challenge in metabolomics and lipidomics studies where analytical methods are typically biased towards molecules from certain chemical classes [58, 93, 94]. For multi-omics integration, ORA is typically performed separately on each omics level. By mapping omics, such as transcriptomics, proteomics, or epigenomics, back to the gene level, multiple omics types can be integrated alongside metabolomics data. The resulting p-values are combined into a joint enrichment p-value for each annotation term using Fisher's method [110] or Stouffer's method (unweighted [111] or weighted [112]) as implemented e.g. in the web-resources PaintOmics3 [74], Integrated Molecular Pathway-Level Analysis (IMPaLA) [113], and MetaboAnalyst [90, 114]. MetaboAnalyst additionally offers an integrative overrepresentation analysis in which both genes and metabolites are queried together by using annotation terms such as metabolic pathways from KEGG to define sets. A drawback of ORA is that it only considers the subset of measured entities that, for example, showed a significant change in levels between conditions. This makes it sensitive to the chosen significance cutoff, or any other inclusion criterion, that was used to determine the input set of biological entities. At the same time, ORA neglects information on the extent of change (e.g., measured through fold change) between conditions [43].

Functional set enrichment analysis (FSEA) is another set-based enrichment method that addresses these ORA-associated limitations. It was originally developed for the analysis of transcriptomics data in Gene set enrichment analysis (GSEA) [115], but has also been implemented for metabolites (Metabolite Set Enrichment Analysis or MSEA) [109] and lipids (LION/web) [116]. In contrast to ORA, these methods test all measured entities, not just a defined subset, and take into account their quantitative measurements. This enables the identification of annotation terms where only a few entities are significantly changed or where many entities are changed slightly but consistently [109]. Similar to ORA, an integrative analysis of several omics datasets is achieved by calculating a joint p-value from the individual single-omics analyses. This is, for example, implemented in the web resource IMPaLA which uses Wilcoxon's signed-rank test to perform FSEA using pathway annotations taken from eleven public databases [113].

The central limitation of both FSEA and ORA is that they are naturally restricted to entities that have been previously annotated. To this end, de novo enrichment methods, such as KeyPathwayMiner [117, 118], have been proposed. These methods enable the discovery of uncharacterized pathways by extracting connected subnetworks with a high number of differentially regulated entities from predefined biological networks (e.g., knowledge-based metabolic networks or data-driven correlation networks) [119]. This framework is theoretically applicable to multi-omics data by using pathway annotations or ontologies that include multiple layers of omics. So far, they have been predominantly used in gene-centric studies. For example, Soerensen et al. [120] demonstrated the benefits of using both GSEA and KeyPathwayMiner in an integrative enrichment analysis of genes associated with cog-

nition in both epigenome-wide and transcriptome-wide association analysis. GSEA was able to replicate findings from previous studies by identifying a broad spectrum of enriched biological processes including gene sets involved in neurological functioning and cell cycle control. The use of *de novo* enrichment identified subnetworks of dysregulated entities that included genes not implicated by GSEA such as Ras And Rab Interactor 3 (*RIN3*) and Ataxin 2 (*ATXN2*). Interestingly, this approach also implicated amyloid beta precursor protein (*APP*) and the nuclear respiratory factor 1 (*NRF1*), two genes with functions relevant to cognitive health, that were not differentially methylated and expressed in this analysis.

Constraint-based metabolic modeling

Constraint-based metabolic models (CBMMs) enable the *in-silico* description and prediction of possible metabolic steady states by mathematically representing metabolic reactions in a stoichiometric matrix [121]. The stoichiometric coefficients of these reactions are used to constrain the flow of metabolites through the system, ensuring that, at steady state, the mass of any compound that is being produced must equal the total amount of what was consumed (flux balance) [122]. Genome-wide metabolic models (GEMs), such as Recon3D, are typically constructed in a bottom-up approach [123] using genome annotations to automatically build a draft that contains all enzymatic reactions predicted to be available for an organism considering the proteins encoded in its sequenced genome. This draft is then refined through manual curation and constraint-based modeling (e.g. to identify and fill gaps in the reconstructed metabolic network) [124].

In the context of multi-omics integration, GEMs present comprehensive metabolic networks that can be used to link the results from single-omics analyses to other layers of biological information by projecting high-throughput data (e.g. transcriptomics, proteomics, or metabolomics data) onto the network [125], analogously as in Section 2.3.1. For instance, GEMs can be used as the underlying biological network in *de novo* pathway enrichment analysis to identify subnetworks that are significantly enriched with dysregulated entities [126].

Furthermore, generic GEM drafts can be contextualized to a specific condition, tissue, or individual by imposing additional layers of constraints that are inferred from experimental omics data [127, 128]. Constraint-Based Reconstruction and Analysis (COBRA) [124, 129] is a popular framework that has implemented multiple methods for the integration of omics data, including time-course metabolomics data [130] and transcriptomics and proteomics data [131, 132]. Contextualized GEMs provide novel opportunities for metabolic engineering, drug target identification, and personalized therapies [125, 127, 133]. For example, Agren et al. [134] used proteomics data of hepatocellular carcinoma patients to construct personalized, cell-specific GEMs for the prediction of antimetabolites (drugs that are structural analogs of metabolites) that can prevent tumor growth. The authors identified nearly

150 antimetabolites, one-third of which were specific to individual patients. Despite the small sample size (n=6) and restricting modeling to cellular effects, this study highlights the potential of refining GEMs using experimental omics data for personalized therapies. The recent emergence of whole-body metabolism reconstructions [135] that currently model human metabolism across 20 organs are expected to further advance this important field.

2.3.2 Data-driven approaches

Data-driven, multi-omics integration approaches use statistical models and machine learning techniques to infer relationships between and within layers of multi-omics data and in some cases a phenotype of interest. Without taking known biological relationships or annotations into account, most approaches rely on the analysis of correlation structures within the data itself. For multi-omics studies focusing on a specific disease or phenotype, common applications of data-driven methods include the training of predictors and classifiers and the identification of multivariate biomarker candidates. Independent of a specific phenotype of interest, the unbiased analysis of relationships between and within omics layers using data-driven approaches enables a global perspective on interactions between biological entities. Using sufficiently large datasets, this approach has the potential to uncover unknown relationships (e.g. not represented in knowledge bases) and to characterize entities with unknown functions.

The following will introduce a selection of step-wise and simultaneous integration strategies and highlight their application in metabolomics and lipidomics studies. A list of multi-omics integration methods and frameworks is provided in Table 2.2.

Step-wise integration

Step-wise strategies integrate datasets in a sequential manner. Here, individual omics layers are typically analyzed separately or in specific (lower-order) combinations. In subsequent steps, the results from these analyses are integrated into a common framework. The following section will introduce ensemble approaches that are suitable for studying a specific phenotype or outcome of interest, as well as pairwise association-based strategies that enable systematic and large-scale integration without necessarily focusing on a specific disease or phenotype.

Ensemble integration strategies apply multivariate classification or prediction methods, such as k-nearest neighbors [136] or Elastic Net [44] to each dataset individually and then combine the ensemble of results using, e.g., majority voting schemes or stacked generalization to boost performance [137]. Although each dataset is modeled separately, these types of methods require omics data that was collected from the same samples as the predictions are ultimately combined in a global model. For example, Ghaemi et al. [44] built a multivariate model predictive of gestational age on samples from 17 pregnant women at three time points during pregnancy. The datasets included measurements from the

immunome, transcriptome, microbiome, proteome, and metabolome. Using the Elastic Net algorithm, the authors built multiple predictors (one for each omics dataset) and subsequently used their predictions as input for a final model. This stacked generalization strategy was able to significantly increase performance and ablation analysis [138] gave insights into the respective contribution of each dataset. Furthermore, subsequent analysis of the top predictive features of each individual model enabled the formulation of multi-omics-informed hypotheses. Among other findings, the authors identified a strong correlation between pregnanolone sulfate and NF-kB signaling in myeloid dendritic cells and regulatory T cells, highlighting a potential regulatory role of this endogenous steroid in the functioning of specific immune cells during pregnancy.

Training the base models in ensemble approaches in an isolated fashion, i.e., on each omics dataset separately, has several consequences. On the one hand, interdependencies between variables of different omics datasets are not fully taken into account such that some cross-omics interactions might be missed. On the other hand, the independence of the base models prevents datasets with a large number of variables from dominating the analysis.

The integration of *pairwise association results* is another step-wise integration strategy. In contrast to ensemble integration, this approach enables the global analysis of relationships between multiple omics layers by large-scale integration of data from multiple sources. A popular approach, which is centered around the concept of genetic variation as a driver of inter-individual variability, is quantitative trait loci (QTL)-based integration [16]. The basis for this integration technique are so-called QTLs [139]. QTLs are genetic markers (e.g., single nucleotide polymorphisms) that are significantly associated with the variation of quantitative molecular traits (e.g., the transcription level of a particular gene) [140]. They are identified in genome-wide association studies (GWAS) that make use of genome-wide genotypes of a large population of individuals that are tested in univariate analyses for association with molecular traits [140, 141, 142]. Besides QTLs of expression levels of genes (eQTLs) [143, 144], major examples of investigated traits include abundances of proteins (pQTLs) [145, 146] or concentrations of metabolites (mQTLs) [23, 147]. For instance, Shin et al. [25] investigated genetic influences on more than 400 human blood metabolites in close to 8,000 individuals from two population-based cohorts. The result is a comprehensive atlas that links genetic variants in 145 loci to biochemical readouts, cataloging mQTLs influencing a wide variety of metabolic pathways.

After association analysis, variant annotation [148] or co-localization analysis [149, 150] is used to functionally interlink entities from different omics by identifying overlapping QTLs (Figure 2.2 C). This can be done on a genome-wide scale and with QTLs that have been identified in different studies or cohorts. QTL-based integration has been successfully applied in studies predicting the functional consequences of disease-associated variants, which are often located in non-coding regions of the genome [146, 151, 152]. For example, Chen et al. [152] systematically overlapped variants associated with autoimmune diseases

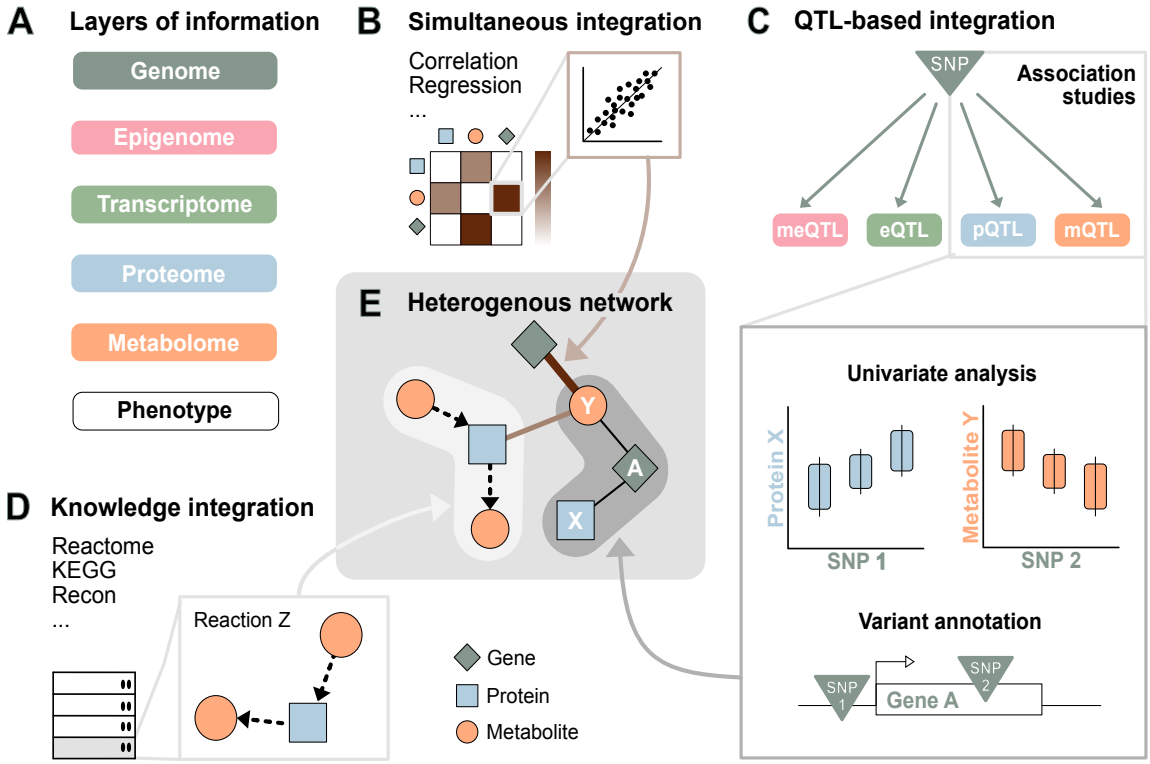


Figure 2.2: Multi-omics integration through composite networks. (**A**) Different layers of a biological system that can be profiled using high-throughput technologies and are frequently integrated in multi-omics studies. (**B**) Simultaneous integration. Correlation structures within and across omics datasets are analyzed using statistical methods. (**C**) QTL-based integration. Using the genome as an anchor, QTLs identified in GWAS are overlaid to establish links between different omics layers. (**D**) Knowledge integration. External information from metabolic databases or scientific literature is used to establish relationships between biological entities. (**E**) Composite networks. By merging the networks inferred in (B-D) on common entities, comprehensive multi-omics catalogues can be constructed. These heterogeneous networks can be mined in post-integration analysis using established graph algorithms. Figure and caption reprinted from Wörheide et al. [10].

with eQTLs as well as DNA methylation (meQTLs), RNA splicing (sQTLs) and histone modification (hQTLs) QTLs to identify cell-specific regulatory effects. Similarly, Suhre et al. [146] demonstrated the power of connecting GWAS-identified risk-variants to disease endpoints via blood proteome-derived pQTLs that overlapped with meQTLs, eQTLs, protein glycosylation QTLs, and mQTLs. Among other findings, this approach revealed a potential link between AD and messenger ribonucleic acid (mRNA) splicing through linking protein levels of apolipoprotein E (*APOE*), a gene centrally linked to AD [153], and small ribonucleoprotein F via overlapping QTLs.

Although this integration strategy only takes into account pairwise relationships, it facilitates the large-scale integration of omics datasets from different sources. This is especially valuable in settings where sufficiently large multi-omics studies in the same set of samples are not available. Furthermore, QTL-based integration only requires summary statistics (results of an association study), circumventing data sharing restrictions that may be present on datasets with patient information. Lastly, this approach can inte-

grate results from independent GWAS on the same traits, providing an opportunity to build data confidence by independent replication. Similarly, meta-analysis methods [154] that statistically combine summary statistics from independent association studies on the same traits (e.g., multiple GWAS with metabolic traits) can be used to increase power and reduce false-positive findings. It is important to note that the concept of integrating pairwise-association results is not restricted to using the genome as an anchor but can be centered around any other omics layer, including the metabolome.

Simultaneous integration

Simultaneous integration strategies use all available omics datasets at the same time and integrate the information in a single modeling step. This has the advantage of taking into account correlations between entities within and across omics layers. The following will introduce selected approaches and categorize them into single-block and multi-block strategies. Single-block integration strategies concatenate all available datasets to form one large data matrix (a 'single block') before applying any analysis method without consideration of heterogeneities between omics (e.g. in scale or variance). In contrast, multi-block integration strategies retain and account for the multi-block structure of the data that is defined by the different omics datasets. Both strategies require that full multi-omics profiles are available for the same set of samples/individuals. Some methods enable the imputation of missing single-omics profiles for a moderate amount of samples/individuals in a multi-omics context. These include MI-MFA (Multiple Imputation - Multiple Factor Analysis) [54] that uses hot-deck imputation [155] to replace missing omics vectors with observed values from a similar sample, and Multi-Omics Factor Analysis (MOFA) [156,157], a statistical framework that infers a low-dimensional data representation in form of (hidden) factors [158]. However, although imputation can increase power by extending the set of available observations, imputed values can never accurately represent the 'true' unobserved measurements and should therefore always be interpreted with caution.

In order to integrate different omics datasets, *single-block integration strategies* simply concatenate the different data matrices into one large data matrix before applying a statistical analysis method. This enables the direct application of methods that are typically applied to single-omics datasets for tasks such as clustering (e.g., K-means clustering [67]), classification, and regression (e.g., Random Forest [159], LASSO regression [160]) or projection (Partial Least Squares Discriminant Analysis (PLS-DA)) [161, 162]). Correlation-based strategies are another popular class of single-block methods, which aim at quantifying the relationships between biological entities by iteratively applying an association measure, such as Pearson's Correlation Coefficient, to all pairwise combinations of the variables (measured biological entities). However, simple correlation measures cannot distinguish between direct and indirect effects [163]. For example, associations between mRNA levels are quite frequently mediated by transcriptional co-regulation at the gene level [164]. These confounded associations lead to a drastically inflated number of edges, resulting

in dense networks with limited interpretability [49, 164, 165]. Gaussian Graphical Models (GGM) [166] circumvent this problem by estimating full-order partial correlation coefficients, i.e., pairwise correlations between variables corrected against all other variables. This measure of conditional independence has been valuable to infer pathway relationships from single omics datasets [63, 167, 168]. However, GGMs assume multivariate normally distributed data and multi-omics datasets often include variables with different distributions, such as phenotypic data on gender or disease subtype [49, 165]. An extension to GGMs that addresses this issue are Mixed Graphical Models (MGM) [169,170,171], which can incorporate datasets with mixed distributions (e.g., continuous, discrete, and count variables) [165]. For example, Zierer et al. [49] inferred an MGM from a multi-omics dataset collected from the same individuals, including data on epigenomics, transcriptomics, glycomics, metabolomics, and phenotypic data. The authors used a Graphical Random Forest [171] method for the integration of 144 preselected features and explored the molecular underpinnings of age-related diseases and co-morbidities. They identified seven network modules that reflect distinct aspects of aging, such as lung function, bone density, and renal function. Furthermore, they found that these modules are connected by distinct hubs, highlighting central molecules and potentially linked mechanisms that may drive co-morbidities, such as urate that connects renal disease with body composition and obesity.

Single-block integration ignores heterogeneities between data types which can lead to severe bias and other complications [39, 41, 172]. For example, metabolomics and transcriptomics data are generated by fundamentally different analytical technologies. This leads to values with different scale and variance as well as different noise distributions [6,173]. When clustering such datasets, the entities within a particular omics type will predominantly cluster together, reflecting intra-, instead of inter-, omics relationships [27, 44, 49, 173]. Similarly, variance maximizing approaches, such as PCA and Partial Least Squares (PLS), will capture these technical differences in their first component [173]. Additionally, the number of variables in each single omics dataset will in most cases be substantially different: a state-of-the-art genomics analysis will provide information on millions of genetic variants, transcriptomics measures tens of thousands of mRNAs, and proteomics and metabolomics technologies usually measure molecules in the range of thousands of molecules [6]. Analyzing such datasets simultaneously without accounting for the diverging numbers of features will introduce bias, as the data type with the most features will drive the results [174].

To circumvent this problem and ensure that every dataset has equal weight, variables can be scaled to unit variance with subsequent block scaling [173] by using, for example, the inverse number of variables in the respective dataset ('block') to scale each variable. This was implemented in Multiple Factor Analysis [174,175], where data blocks are normalized prior to concatenation by using the inverse of the first squared singular value of a PCA on each data block as weight. However, different methods for variable scaling and block

scaling can significantly influence the outcomes [173]. General caution is advised when concatenating datasets from different sources and special care should be taken to identify an integration method that combines and scales data appropriately [16, 173].

The need to account for heterogeneities between multi-omics datasets has led to the emergence of *multi-block integration strategies* that can take the block structure, i.e., groups of omics variables from different sources, into account [176]. Multi-block methods simultaneously model multiple data matrices and provide insights into the relationships between omics (blocks). Many of these approaches are extensions of established multivariate methods, such as PLS. Examples include O2PLS [177, 178] for the integration of two omics datasets and Multiple-Block Orthogonal Projections to Latent Structures (OnPLS) [179,180,181] for the integration of more than two omics datasets. OnPLS decomposes data from multiple omics data matrices into global, local, and unique levels of variation [181]. Reinke et al. [182] demonstrated the potential of this approach using a small subset (n=22) of individuals from an asthma cohort. Here, six blocks of data - transcriptomics, metabolomics, three targeted assays (on sphingolipids, oxylipins, and fatty acids), and clinical variables - were integrated using OnPLS. Subsequent variable selection and visualization gave insights into cross-omics interactions, for example, by identifying a potential link between transcript levels of *ATP6V1G1*, a gene that has been associated with osteoporosis, and multiple metabolites that are dysregulated by inhaled corticoid steroids.

Other popular multi-block integration strategies include unsupervised methods such as regularized generalized canonical correlation analysis (RGCCA) and sparse generalized canonical correlation (SGCCA) [183], as well as the supervised framework Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO). DIABLO [47] is a multivariate classification method that extends SGCCA to a supervised analysis and prediction framework. It can identify key omics variables that drive the discrimination between phenotypic groups of interest and simultaneously builds a predictive model to classify new data [45, 48, 184, 185, 186]. For example, Qui et al. [48] integrated genomic, transcriptomic, epigenomic, and metabolomic datasets from patients with high and low bone mineral density. Using DIABLO, they identified a multi-omics biomarker panel for osteoporosis that includes 74 differentially expressed genes, 75 differentially methylated CpG sites, and 23 differentially abundant metabolites. To gain further mechanistic insights into underlying disease mechanisms, the authors conducted a targeted QTL-based analysis in combination with Mendelian randomization. They were able to identify five biomarkers (*ADRA2A*, *FADS2*, *FMN1*, *RABL2A*,*SPRY1*) with a causal effect on levels of bone mineral density. DIABLO and various other projection-based integration methods are implemented in the R package mixOmics [172] which is focused on data exploration, dimensionality reduction, and visualization of multi-omics data.

Simultaneous integration strategies have been applied by relatively few studies so far, with mostly small numbers of samples/individuals. This is most likely due to the lack

of larger available multi-omics datasets. Nevertheless, simultaneous integration, and especially multi-block methods, are powerful tools that have the potential to fully exploit multi-omics data in integrative analyses.

2.3.3 Composite network approaches

Composite networks aim at capturing relationships between omics layers in heterogeneous networks by merging information from different knowledge-driven and/or data-driven sources (Figure 2.2). This step-wise integration strategy is gaining increasing popularity due to its scalability and versatile applicability. In order to construct a composite network, the information from each knowledge-based (e.g., STRING, KEGG) or data-driven (e.g., correlation-based) component is stored and interconnected in accessible network structures (graphs) that are merged by overlaying common biological entities (Figure 2.2 B-E). This can be accomplished by simple concatenation of the respective underlying edge lists, provided that there is some degree of overlap between the datasets and/or resources. The resulting network consists of nodes (biological entities such as genes, proteins, and metabolites) connected by edges that model pairwise functional, biochemical or physical relationships [187]. Composite networks are per se not bound to a specific phenotype or disease of interest. Once built, they provide a comprehensive catalogue of inter- and intra-omics relationships that can be explored in post-integration analyses to identify and prioritize relevant entities in the neighborhood of e.g. disease-associated genes within the network or to predict novel associations.

Composite networks can be built in a knowledge-based, data-driven, or hybrid fashion. While knowledge-based integration allows the large-scale analysis of vast amounts of published information without requiring additional omics experiments [51], this approach is restricted to entities that have been annotated. Data-driven composite networks merge inferred information from experimental multi-omics data and, in contrast, can naturally only include the biological entities measured by the respective omics technology. By combining these two approaches, for example, by extending data-driven networks (e.g. built through QTL-based integration described in Section 2.3.2) with knowledge-based relationships (e.g., gene-transcript-protein or drug-drug targets relations), it is possible to construct comprehensive multi-layered resources that facilitate the unbiased generation and exploration of multi-omics hypotheses. HENA [188], a heterogeneous network-based dataset for AD, is a recent example of this. Sügis et al. integrated data relating to AD, including GWAS results, protein-protein interaction, and gene co-expression networks, from public knowledge databases and experimental datasets. The resulting gene-centric network was subsequently analyzed using graph convolutional networks to identify disease-related genes, highlighting one of the many potential applications of composite networks. Future frameworks that additionally include metabolite data will provide even more comprehensive models for studying molecular mechanisms implicated in AD.

Although conceptually simple, the construction of composite networks is complicated at large due to the discussed challenges of ID mapping and compound identification (see Section 2.3.1), as well as differing data formats between resources, and considerations regarding statistical cut-offs and weighting of information types. Furthermore, the post-integration analysis of these large and highly complex networks is not straightforward and requires sophisticated algorithms (further discussed in Section 2.4). Consequently, databases and frameworks that provide access to composite networks are attracting growing interest, such as ConsensusPathDB [104, 189] and omicsNet [102, 190].

2.4 Post-integration analysis, visualization and interpretation

Post-integration analysis of inferred networks or multi-omics features through manual inspection or computational algorithms is key to gaining biologically relevant insights and fully exploiting the potential of multi-omics datasets. So far, a limiting factor has often been the ability to represent, comprehend and reproduce highly complex and multifactorial relationships across multiple biological domains [191].

For studies that are driven by a clear research question, interpretation can be straightforward. For instance, when building a predictor for a specific phenotype of interest, integration methods such as DIABLO (Section 2.3.2) result in a subset of interesting (in a statistical sense, e.g. most predictive, most significant) biological entities. This set of variables can then be subjected to downstream analyses to gain further functional insights or to investigate causality (e.g. via Mendelian randomization). Global integration efforts, on the other hand, enable exploratory analysis by systematically cataloging biological entities and their interactions without focusing on a specific phenotype or disease. Here, post-integration analysis through computational algorithms provides tools to identify patterns in the data and pinpoint interesting entities.

To this end, networks provide a flexible and intuitive mathematical framework to represent, visualize, and analyze these complex relationships [192]. Various techniques have been developed that facilitate the visual representation and exploration of networks in a human-comprehensible form by arranging nodes and edges in specific layouts. For example, by grouping nodes together that are highly connected, modular patterns in the data become more visible [192]. However, with growing complexity and size, networks can quickly become very dense and difficult to comprehend [193]. Alternative representations of large networks, such as structural summary [194] or axis-based node-link representations [195] have been developed to mitigate these challenges and provide scalable layout alternatives [196].

Table 2.2: A selection of multi-omics data integration frameworks and methods. Table and caption reprinted from Wörheide et al. [10].

	Matching samples	Integration strategy	Implementation	Reference	Description
Knowledge-based					
IMPaLA	no	enrichment	online	[113]	Integrated Molecular Pathway Level Analysis (IMPaLA) enables joint pathway analysis.
COBRA	-	constraint-based modelling	MATLAB Python Julia	[75, 124]	The COnstraint-Based Reconstruction and Analysis (COBRA) Toolbox.
PathMe	-	composite network	online Python	[105]	Integrates KEGG, Reactome and WikiPathways into a unified abstraction.
Data-driven					
KeyPathwayMiner	no	*de novo* enrichment	online Cytoscape	[117, 118]	Extracts all maximal connected sub-networks which enriched for dysregulated entities.
MI-MFA	partially	imputation/ ensemble	R code in supplementary	[197]	Uses multiple imputation (MI) to enable the application of multiple factor analysis (MFA) to multi-omics data with partially missing single-omics profiles.
MOFA	partially	imputation	R Python	[156, 157]	Unsupervised integration framework that infers a low-dimensional data representation and enables the imputation of missing omics profiles.
causalMGM	yes	single-block	online	[198]	Learns a causal (i.e., directed) graph using variable selection with subsequent application of a mixed graphical model (MGM) PC-Stable algorithm.
omicade4	yes	single-block	R	[197]	Projection-based method that performs multiple co-inertia analysis.
xMWAS	yes	single-block	online R	[199]	Uses (sparse) Partial Least Squares regression to perform pairwise correlation analyses and build a heterogeneous network.
mixOmics	yes	multi-block	R	[172]	Collection of unsupervised and supervised multivariate methods, including sparse generalized canonical correlation (SGCCA) and Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO).
OnPLS	yes	multi-block	Python	[180, 181]	Projection-based integration method that decomposes global, local and unique levels of variation.

In addition to providing intuitive visualization, networks enable the application of a rich toolbox of established graph algorithms to explore multi-omics networks and extract relevant information in an automated manner [200]. For example, multi-layer networks represent a promising mathematical framework, where layers of nodes (e.g., genes, proteins, metabolites) are connected by different edge types with varying degrees of connectivity (e.g., gene co-expression, trait association, and protein co-abundance) [201, 202]. Research fields such as graph theory and network science have developed various algorithms that can be applied to such heterogeneous networks, including random walk [51], module identification [203], or meta-path-based techniques [204]. This enables, for example, the prediction of novel edges [204], the identification of key players [205, 206], or retrieval of interesting subnetworks (modules) [207,208,209]. Furthermore, native graph databases, such as Neo4j, represent an attractive framework for post-integration analysis as they enable the efficient storage and analysis of large amounts of semi-structured, diverse, and highly connected data [7]. An extensive list of network-based multi-omics visualization tools and online resources is provided in Table 2.1.

Even after the successful identification of interesting entities or modules, the downstream functional interpretation and validation of such complex multi-omics findings is not straightforward. Direct replication as an important tool for identifying false positives [16] is often not an option due to the frequently limited availability of comparable and sufficiently powered omics studies. So far, validation of results has therefore often been performed using prior knowledge [191] to provide functional evidence, for example, through set-based enrichment (Section 2.3.1). However, with growing numbers of large-scale studies and efforts towards standardizing and indexing datasets across sources, such as the Omics Discovery Index (OmicsDI) [210,211], data-driven replication will become increasingly feasible in the future. Beyond that, it is often not possible to describe every finding from a multi-omics study in detail as results can be very complex and numerous. This consequently leads to biased or selective reporting of outcomes that are published [191]. To this end, the sharing of all results in easily accessible data repositories, such as NDEx [101], or dedicated supplemental web servers [25, 29, 146, 212], is becoming more popular as it enables the re-use of multi-omics results for further exploration or replication by other researchers.

2.5 Current trends and future perspectives

As highlighted in this Chapter, various multi-omics integration strategies exist. Developments in research fields such as computer vision and natural language processing offer promising new directions for the unbiased integration of high-dimensional data. Recently, these fields have been transformed by the use of deep learning techniques, such as deep neural networks, which can handle vast amounts of data and are able to discover highly complex and relevant features [213, 214]. In deep learning, multiple hidden layers enable the learning of new, highly complex data representations [213]. Furthermore, flexible architectures allow models to be tailored to many different problem domains, providing exciting new possibilities also for multi-omics integration studies [215, 216]. For example, variational autoencoders (VAEs) [217] are popular representation learning methods that have been proposed for non-linear dimensionality reduction, unsupervised clustering, and denoising of datasets [218,219]. They can be used to encode input data (e.g., different omics datasets) into a low-dimensional embedding, effectively integrating different omics types into a new latent representation [220]. A major limitation of deep learning algorithms, so far, has been their need for vast amounts of high-quality data and the complicated interpretation of model features [214, 216, 221]. However, the increasing availability of large multi-omics datasets and development of interpretable deep learning methods will enable more and more deep learning applications in the future [213, 221].

Besides algorithmic innovations, the ongoing advances in analytical technologies will also provide novel opportunities and challenges for integrative studies. For example, spatial omics profiling has received increasing attention in the past few years due to the advent of high-resolution technologies to generate data in fine-grained spatial resolution. This is

particularly interesting for the cancer field, where there is increasing evidence that the tumor microenvironment, i.e., the collection of all stromal cells surrounding and supporting the tumor cells, plays a major role in prognosis and therapy [222]. For metabolomics, modern Matrix Assisted Laser Desorption Ionization (MALDI)-imaging mass spectrometry instruments can acquire metabolite profiles at almost single-cell resolution [223]. This rich new type of data, composed of metabolites, samples, and two or more spatial dimensions, also requires innovative approaches for data processing, integration, and analysis. For example, single-cell metabolic profiles can be assigned and analyzed using the 'SpaceM' method, which performs the interpolation of spatial measurement patterns onto microscopy images [224]. Similarly, new technologies and the corresponding computational methods allow for high-resolution protein profiling, e.g., using mass cytometry time of flight (CyTOF) instruments [225], and spatial transcriptomics data can be obtained by a growing number of sequencing and microarray-based techniques [226]. Future applications, where tissue samples or entire organs are analyzed in a sequential fashion with a combination of these techniques to generate spatial multi-omics datasets, promise unprecedented insights into the deep molecular biology of the systems under study.

2.6 Summary

The generation of vast amounts of biological data has provided exciting new opportunities to gain a systems view on molecular wirings across regulatory layers that define health and disease. However, the heterogeneous and high-dimensional nature of multi-omics datasets in combination with differing study objectives and data scenarios make the appropriate data integration strategy a case-by-case choice.

While knowledge-based strategies can guide integrative analysis by harnessing a large body of manually and experimentally validated information from databases and scientific literature, it is restricted to known or previously characterized biological entities and is not applicable to molecules with unknown functions or identities. Data-driven methods, on the other hand, use statistical methods, such as correlation or association analysis to infer relationships between omics layers. Although this can be prone to the identification of spurious associations and success heavily depends on correctly preprocessed, high-quality data, data-driven integration has the potential to discover novel as well as condition-specific interactions. In particular, multi-block integration methods that can simultaneously analyze datasets while taking into account inter-omics heterogeneity show exciting potential to fully exploit multi-omics datasets. To leverage the advantages of both approaches, network-based hybrid integration methods have emerged that enable the combination of knowledge-based and data-driven data integration. This facilitates the generation of highly complex multi-omics interaction catalogues that can be mined in an automated fashion using graph algorithms.

With the increasing availability of larger, high-quality datasets paralleled by the development of new omics technologies, the demand for powerful data analysis tools and standardized integration frameworks will continue to grow. The integrative analysis of these multi-omics data, enabled by publishing data in centralized data repositories adhering to the FAIR Principles (Findable, Accessible, Interoperable, and Reusable) [227], will finally allow researchers to promote the usability and reproducibility of their work and has the potential for achieving substantial advances in biomedical research and health care.

Alzheimer's disease (AD) 3

"Alles in allem genommen haben wir hier offenbar einen eigenartigen Krankheitsprozeß vor uns." On the whole, it is evident that we are dealing with a peculiar disease process.

— Alois Alzheimer, *Allgemeine Zeitschrift fur Psychiatrie und Psychisch-gerichtliche Medizin.* 1907

When Alois Alzheimer (1864-1915) first described the medical case of Auguste Deter (1850-1906) at a psychiatry conference in Tubingen in 1906, he was probably not aware of the far-reaching implications his discovery would have. Auguste Deter was a 51-year-old patient who was assessed in 1901 by Alois Alzheimer, presenting with unusual symptoms, including memory loss, sleeplessness and other psychological changes [228]. After her death, Dr. Alzheimer and his colleagues examined her brain using slide preparations stained with the Bielschovsky silver method, discovering unusual changes in the neurofibrils of her brain cells. He described his histopathological findings in a later publication as [229]:

> "[...] inside an apparently normal-looking cell, one or more single fibers could be observed that became prominent through their striking thickness and specific impregnability. [...] At a more advanced stage, many fibrils arranged parallel [...] they accumulated forming dense bundles. [...] Dispersed over the entire cortex, and in large numbers especially in the upper layers, miliary foci could be found which represented the sites of deposition of a peculiar substance in the cerebral cortex."

Although his findings were not immediately met with recognition, the discovered brain lesions, namely amyloid-β (Aβ) plaques, and neurofibrillary tau tangles would later become the defining pathologic features of a devastating new neurodegenerative disease. The disease was first referred to as Alzheimer's disease (AD) in 1910 by Emil Kraepelin, a German psychiatrist and colleague of Dr. Alzheimer and Auguste Deter is considered to be the first described AD case [228].

Today AD is recognized as a devastating progressive, neurodegenerative disorder that is estimated to affect a total of 6.5 million people of age 65 and older in the United States alone. This number is projected to rise to 13.8 million by 2060 [230, 231] and poses a significant financial and social burden on society [232]. AD is the most common cause of dementia in older adults, accounting for 60% to 80% of dementia cases [232]. Its pathological hallmarks include the presence of extracellular deposits of Aβ (amyloid plaques) and intracellular neurofibrillary tangles of hyper-phosphorylated tau aggregates in the brain, leading to synapse loss, brain atrophy, and ultimately death (Figure 3.1). The aggregation of these neuropathological changes (further described in Section 3.1) can start decades before any clinical symptoms are observable. The symptoms of progressing AD include episodic memory impairment, executive dysfunction, and changes in behavior and personality. Furthermore, clinicopathological studies have shown that older individuals living with AD often present mixed pathologies, meaning they have one or more neurodegenerative and cerebrovascular disease pathologies present [233]. Only a small percentage of individuals are believed to have 'pure' AD, making it difficult to study the disease in complete isolation from other age-related disorders and increasing the heterogeneity of clinical presentations and trajectories [232, 233, 234].

The exact causes of AD are not fully understood, but a combination of genetic and environmental factors are thought to contribute to the onset and progression of the disease. A small percentage of AD cases can be attributed to autosomal dominant mutations in the amyloid beta precursor protein (*APP*), presenilin 1 (*PSEN1*), and presenilin 2 (*PSEN2*) genes, affecting *APP* proteolysis and synthesis [235]. Individuals carrying such a mutation are almost certain to develop AD before they reach the age of 65, which is why this form of AD is also referred to as early-onset Alzheimer's disease (EOAD) [232]. Sporadic AD on the other hand is the most common form of AD and typically has an onset in individuals over 65 years of age. This late-onset Alzheimer's disease (LOAD) is not caused by a single genetic mutation, but rather a complex interplay of genetic and environmental factors. Risk factors for LOAD include:

Age. Although AD is not considered a normal part of aging, age represents the strongest risk factor for developing LOAD, as the percentage of individuals living with AD increases substantially with age. For example, while the prevalence (in the United States) of clinical AD is 5.3% among adults 65 to 74 years old, it is increased to 34.6% when looking at adults 85 years or older [230].

Genetics. LOAD is a complex, polygenic disorder with heritability estimates ranging between 60 to 80% [236]. Multiple genetic risk factors for LOAD have been identified with the strongest risk exerted by the APOE*ε4 allele of apolipoprotein E (*APOE*) [237]. *APOE* has three common forms, the least common APOE*ε2 (reduces AD risk), the most common form APOE*ε3 (no effect on AD risk) and the risk increasing APOE*ε4. Individuals that inherit a single copy of APOE*ε4 have an approximately 3- to 4-fold risk of

developing AD while two copies increase the risk to 9- to 15-fold [238]. *APOE* is involved in the re-distribution of cholesterol and other lipids to neurons and is thought to drive amyloid pathology [238]. To date, more than 40 loci have been linked to AD risk through genome-wide association studies (GWAS), including cerebrospinal fluid clusterin (*CLU*), *CR1* and *PICALM* [239]. However, most studies have been carried out in individuals of European ancestry and studies have shown differing allele frequencies, risk profiles, and disease prevalence among different ethnic groups [230, 240, 241].

Sex. Female sex is considered a major risk factor of AD. It has been estimated that at 45 years of age, the lifetime risk of AD in women may nearly be double that of men [232, 242] and women with Mild Cognitive Impairment (MCI) may experience faster progression to AD dementia than men [243]. Although women live longer than men and age is the greatest risk factor of AD, longevity seems to only partially explain the observed sex differences in AD susceptibility and severity [244]. Sex-specific effects and interactions of risk factors have been proposed as contributing mechanisms. For example, a recent study showed sex- and APOE*ε4 status-dependant alterations of the human serum metabolome pointing to greater mitochondrial impairment in APOE*ε4 carrying females than males [245].

Other risk factors. Epidemiological studies have linked cardiometabolic risk factors, such as hypertension [246], midlife obesity [247], and type 2 diabetes [248] to an increased risk of AD and dementia, suggesting underlying common disease mechanisms, such as chronic inflammation, mitochondrial failure or insulin resistance [233, 234]. Additional modifiable lifestyle factors that can lead to an increased AD risk include smoking, physical inactivity, and social isolation [249], while cognitive reserve, intellectual and physical activity and a healthy diet have been proposed as protective factors [250].

This thesis is focused on the late-onset form of the disease and in the following, we will use LOAD and AD interchangeably.

3.1 Pathogenesis of AD

AD is a complex disorder that results from various changes in the brain. Amyloid and tau proteins, which form the pathological hallmarks of the disease, Aβ plaques, and neurofibrillary tangles, respectively, have long been assumed to play a significant role in disease pathogenesis. The Amyloid hypothesis suggests that the aggregation of extracellular Aβ in the brain triggers a chain reaction that leads to intraneuronal tau protein accumulation and eventually neurodegeneration [252, 253]. It is supported by evidence from familial EOAD, where mutations in *APP*, the precursor of Aβ peptides, and mutations in *PSEN1* and *PSEN2*, parts of the protein complex that cleaves *APP*, affect Aβ cleavage and aggregation. Furthermore, individuals with Down's syndrome (trisomy 21), who carry an extra copy of the *APP* gene (located on chromosome 21) and produce higher levels of Aβ throughout their life, develop typical AD neuropathology [252]. Taken together, the

evidence for a causative role of Aβ pathology in AD is compelling. However, high failure rates of amyloid targeting drugs, insights from genetic studies, and differences in the spatiotemporal progression of amyloid and tau pathology in the brain, have given rise to an alternative hypothesis, such as the dual-cascade hypothesis [254, 255, 256, 257]. Here, Aβ and tau pathology exist as correlated but independent cellular pathways in AD [257].

The following will provide a short overview of selected biological pathways that have been associated with LOAD. However, it is important to note that the specific temporal ordering and interrelationships of the individual processes are not well understood.

3.1.1 Amyloid-β pathology

Aβ peptides are produced by the sequential proteolysis of *APP*, a transmembrane protein that is enriched in neuronal synapses. This process is also known as the amyloidogenic pathway and is mediated by beta-site amyloid precursor protein–cleaving enzyme 1 (*BACE-1*) and γ-secretase [258]. *APP* can also be cleaved by α-secretase, preventing the production of Aβ. This is known as the non-amyloidogenic pathway. The produced Aβ is secreted into the extracellular space as a monomer and is typically 36 to 43 amino acids long. Thereby, $A\beta_{40}$ is more prevalent than the aggregation-prone $A\beta_{42}$, which has been centrally linked to AD pathogenesis [258, 259]. Extracellular Aβ peptides can spontaneously self-aggregate into oligomers, fibrils, and eventually plaques. In AD, the failure of Aβ clearance systems and/or shift to the synthesis of longer, more aggregation-prone Aβ species are thought to

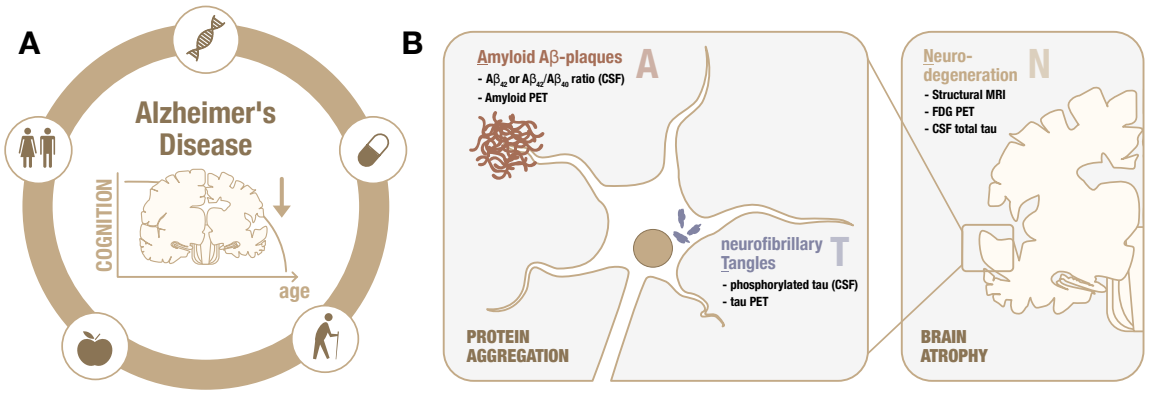


Figure 3.1: Alzheimer's disease (AD). (**A**) AD is a neurodegenerative disorder caused by the accumulation of neuropathological changes in the brain. It is clinically characterized by symptoms, such as cognitive decline, progressively worsening over time. Risk factors for AD include age, genetics, sex, and environmental influences, such as diet or preexisting conditions. (**B**) Neuropathological hallmarks and biomarkers of AD. AD is characterized by the presence of distinct pathological changes in the brain: (A) extracellular amyloid-β (Aβ) plaques and (T) intracellular neurofibrillary tangles that eventually lead to (N) neurodegeneration [251]. These brain lesions can be detected by *in vivo* biomarkers. Biomarkers of fibrillary Aβ deposition are high ligand retention on cortical amyloid positron emission tomography (PET) and low cerebrospinal fluid (CSF) $A\beta_{42}$. Biomarkers of tau pathology are elevated CSF phosphorylated tau (p-tau) and high ligand retention on cortical tau PET. Biomarkers of neurodegeneration are CSF total tau (t-tau), fluorodeoxyglucose (FDG) PET hypometabolism, and brain atrophy seen on structural magnetic resonance imaging (MRI).

increase Aβ levels in the brain. This eventually leads to Aβ aggregation and oligomerization, possibly activating or amplifying a cascade of disease-driving processes, including tau aggregation, neuroinflammation, and synapse loss [252]. Studies suggest that particularly soluble oligomers of $A\beta_{42}$ mediate neurotoxicity in AD [260]. Assessment of the spatiotemporal progression of Aβ pathology using PET neuroimaging techniques have shown that Aβ accumulation starts in association cortices of the brain and spreads from neocortex to allocortex, eventually reaching the cerebellum [257].

3.1.2 Tau pathology

Tau is a microtubule-associated protein that is involved in microtubule stabilization and vesicle transport. It is normally located in the cytoplasm of axons but can also be found in presynaptic and postsynaptic sites. Through post-translational modifications, tau can become hyperphosphorylated and lose its affinity for microtubules, destabilizing the neuronal cytoskeleton and impairing axonal transport [258, 261]. This hyperphosphorylated form of tau self-aggregates into intermediate forms, such as intracellular neurofibrillary tangles. Tau tangles are not exclusive to AD and can also be found in other neurodegenerative disorders (tauopathies). As with Aβ, intermediate aggregates of pathologic tau are cytotoxic and promote neurodegeneration [258]. However, tau pathology correlates more strongly with neurodegeneration and cognitive decline than Aβ pathology [257,262] and has been suggested as a driver for AD-related metabolic changes observed in the brain [263]. Although it has been proposed that Aβ drives the formation of intracellular tau tangles, other factors, such as *APOE* and microglial activation, may also play a role [257, 264, 265]. Tau aggregation typically occurs later than Aβ pathology. It starts in the transentorhinal cortex, then spreads to limbic areas and eventually to the neocortex [257].

3.1.3 Synaptic dysfunction

Progressive brain atrophy, caused by synaptic and neuronal loss, can be observed in the temporal lobes and later in the frontal lobes of AD patients [266]. Synaptic loss is more pronounced than the neuronal loss in progressing AD and is the strongest neuropathological correlate with cognitive decline [258, 266, 267]. Evidence from model organisms of AD implicate both oligomeric Aβ, and pathogenic tau in synaptic dysfunction [268, 269]. Especially Aβ oligomers have been shown to impair synaptic plasticity and reduce dendritic spine density, for example, through the reduction of glutamatergic transmission [270], cholinergic dysfunction [271], and disruption of calcium homeostasis [260].

3.1.4 Lipid metabolism

Dysfunctional lipid metabolism has been implicated in AD onset and progression by multiple lines of evidence, including large-scale genetic [272, 273, 274, 275] and metabolomics studies [19, 19, 263, 276, 277]. GWAS have identified multiple genetic variants that are

associated with AD and occur in genes involved in the regulation of lipid metabolism, including *CLU*, *SORL1* and *ABCA7* [266, 278]. The most prominent example, however, is the risk gene *APOE* which transports cholesterol from astrocytes to neurons in the brain [258]. Cholesterol is an important part of neuronal membranes and, together with sphingomyelins (SMs), is a constituent of highly organized membrane microdomains, so-called lipid rafts. Lipid rafts in the brain facilitate protein and lipid interactions and are involved in various processes, such as signal transduction and regulation of inflammatory processes [279]. Both cholesterol and SMs have been linked to the amyloidogenic processing of *APP* and implicated in AD pathogenesis. For example, specific SMs have been associated with early stages of AD [19] and analysis of the brain transcriptome revealed global dysregulation of the SM pathway [276]. In addition, higher levels of ceramides, the neurotoxic products of SM hydrolysis, are thought to increase *BACE-1* activity, leading to increased Aβ production [277]. Products of cholesterol metabolism have also been associated with AD-traits in multiple studies [263, 280, 281] and it has been suggested that high levels of cholesterol increase the secretion of Aβ through effects on membrane composition and co-localization of *APP* and *BACE-1* [279].

3.1.5 Neuroinflammation and microglia

Increasing evidence suggests that inflammation and the innate immune system in the brain play a significant role in the development and progression of Alzheimer's disease [282]. GWAS have implicated multiple immune response- and microglia-related genes including *TREM2*, *CR1*, *CD33* and *INPP5D* [278, 283]. These findings are supported by transcriptomics and proteomics studies that showed that GWAS candidates are often located in network modules that correlate with tau pathology and cognitive decline, and are enriched for microglial markers [284, 285]. Furthermore, markers of inflammation or increased levels of molecules that are indicative of active inflammatory processes have been observed in AD patients [263, 282, 286]. While various neuroimmune cells have been implicated in AD, microglia have attracted particular interest [287]. Microglia can bind soluble Aβ oligomers and fibrils, and release inflammatory cytokines [282, 287]. As the innate immune cells of the central nervous system (CNS), microglia have also been implicated in Aβ clearance [288]. However, whether microglia play a protective (clearance of Aβ plaques) or harmful (Aβ-induced chronic inflammatory state) role in AD is not clear. Interestingly, a recent single-cell study of AD-transgenic mouse brains has identified a specific subpopulation of beneficial microglia, so-called disease-associated microglias (DAMs), which accumulate around plaques [289]. Their initial activation involves the up-regulation of a set of genes, including *APOE* and *TYROBP* in concert with the down-regulation of homeostatic genes. However, full activation of DAMs, including phagocytic activity, was shown to be *TREM2*-dependent. It has been proposed that impairment of this process may occur in later phases of the disease and lead to a decrease of Aβ phagocytosis and

sustained inflammatory response. Furthermore, this may explain part of the genetic risk conveyed by variants in *TREM2* and *APOE* [266, 290].

3.2 Diagnosis, biomarkers, and treatment

3.2.1 Disease progression and diagnosis

AD encompasses the progression of individuals from unnoticeable pathological brain changes to severe AD dementia. This spectrum of disease severity is also referred to as the Alzheimer's continuum [251]. Using the degree of cognitive impairment, AD can be divided into three main disease phases [291]: (a) a preclinical stage, where neuropathological changes in the brain related to the disease begin to occur, but do not cause clinical symptoms, (b) a prodromal stage (MCI due to AD), where symptoms, such as memory loss, start to manifest but do not substantially interfere with daily activities, and (c) Alzheimer's dementia, which is the final stage of the disease, characterized by symptoms severe enough to affect an individuals ability to live independently. However, it is important to note, that the pace of progression of the disease can vary and not all patients that present AD pathological brain changes or have MCI due to AD will go on to develop Alzheimer's dementia during their lifetime [232, 291].

AD is viewed as a clinical-pathologic entity [251]. As a consequence, AD can be diagnosed as *probable* AD in living individuals through the assessment of patient history, mental status examination, and neuropsychological tests. Frequently used assessments include the 13-item Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) [292] or Clinical Dementia Rating (CDR) [261, 293]. However, MCI and dementia can also be caused by other, non-AD conditions. Therefore, a *definite* AD diagnosis can only be made through an autopsy after death [261]. Postmortem AD diagnosis and staging rely on semi-quantitative measures of neuropathological burden, such as Braak stage [294] for neurofibrillary tangle pathology, Consortium to Establish a Registry for Alzheimer's Disease (CERAD) score [295] for neuritic plaques or consensus of both (e.g., National Institutes on Aging (NIA)-Reagan score [296, 297]). Because these pathological changes in the brain start to accumulate decades before the onset of clinical symptoms, there has been increasing interest in shifting from purely clinicopathological to *in vivo* biomarker-guided classifications of AD [298, 299].

3.2.2 Biomarkers of AD pathology

A biomarker is an objective, quantifiable marker of a biological process that can be measured accurately and reproducibly [300]. Ante-mortem biomarkers for AD neuropathology can, for example, enable the detection of preclinical AD or aid AD drug development by guiding the selection of participants and confirming their diagnosis in clinical trials [301]. The development of neuroimaging and CSF biomarkers have now made it possible to detect

and characterize AD pathologies *in vivo* (Figure 3.1 B), providing exciting new opportunities. For example, physicians can use biomarkers to assist the diagnosis of MCI due to AD to rule out other underlying causes for the observed clinical presentation. Nevertheless, there is still a desperate need for cost-effective, accurate, and non-invasive biomarkers, such as blood tests, that can be routinely used in clinical settings and may eventually pave the road for widespread testing for AD [302]. Blood-based biomarkers for AD diagnosis and prognosis are currently in development and will hopefully see clinical use in the near future [303].

To better capture the heterogeneity of AD in research settings, the NIA and Alzheimer's Association proposed a classification system [251, 299] where AD biomarkers are categorized into three distinct groups; biomarkers for (A)myloid-β plaques, fibrillar (T)au, and (N)eurodegenration or neuronal injury (Figure 3.1 B). Individuals can then be classified into one of eight AT(N) biomarker profiles. For example, the Alzheimer's continuum is defined by an abnormal amyloid biomarker (A+), independent of the status of other biomarkers. Using the proposed terminology, individuals presenting with abnormal A but normal T and (N) (A+/T-/(N)-) show Alzheimer's pathological changes, and AD is defined by neuropathologic or biomarker evidence for both Aβ plaques and pathologic tau deposits (A+/T+/(N)- or A+/T+/(N)+) [251]. Although this purely biological definition of AD has limitations in clinical settings, as it entirely removes a clinical component [302], it provides a unified terminology and classification scheme that can be used in AD research studies to represent individuals across (and beyond) the Alzheimer's continuum.

3.2.3 Pharmacological therapies

AD therapies have had extremely high failure rates (99%) [301]. To date, only seven drugs have been approved by the Food and Drug Administration (FDA) for treating AD. Five of them target neurotransmission and are symptom-based, i.e., do not target the underlying biological causes of disease and can therefore not alter or improve the course of disease [304]. Most of these drugs (tacrine, donepezil, galantamine, and rivastigmine) are cholinesterase inhibitors that can improve or delay the progression of symptoms by preventing the breakdown of neurotransmitters. In contrast, the *N*-methyl-D-aspartate (NMDA) receptor antagonist memantine is believed to reduce the effects of excitotoxic glutamate release. Recently, two potentially disease-modifying treatments have been approved by the FDA through an accelerated approval process. Both drugs, Aducanumab [305] and Lecanemab [306], are monoclonal antibodies targeting Aβ. Aducanumab was given approval based on two terminated phase 3 trials although, in only one of them, the high-dosage subgroup met its primary and secondary endpoints [305]. Lecanemab, on the other hand, was shown to meet all primary and secondary outcomes in a completed phase 3 trial, significantly slowing cognitive decline and reducing amyloid burden in the brain [306]. Aducanumab was also shown to reduce Aβ plaques, but with less benefit on cognition. While

exiting, these developments are not without limitations as both drugs can cause serious side effects [305,306], and the respective companies will have to verify their clinical benefit in post-approval trials [232].

For future clinical trials, recruiting more adequate trial populations [301] (e.g., biomarker-confirmed AD pathology, earlier stages of AD) will be of vital importance [301]. Alternatively, drug repositioning, i.e., the application of available compounds in a novel disease context, has gained increasing attention and poses a promising alternative to *de novo* drug development [307]. Using multi-omics data generated by international and interdisciplinary research efforts, new methods are being implemented to computationally identify and prioritize promising drug candidates for repositioning [277,308]. Nevertheless, AD remains an incurable disease, highlighting our incomplete understanding of the etiology of the disease and the urgent need for new therapeutic approaches.

3.3 Integrative analysis of AD

Integrative analysis across omics holds the promise to provide a more comprehensive understanding of underlying and potentially targetable disease mechanisms. In general, two types of multi-omics AD resources have been developed (i) results explorers and (ii) network-based (cross-disease) datasets.

Results explorers provide researchers with tools to look up multi-omics evidence on a single target level. However, they do not - or only in a limited scope - provide the extended multi-omics context surrounding these entities and/or enable the exploration of molecular links between (manually provided) sets of targets. Current results explorers include:

- **Agora** [309] – is an NIA funded results explorer that provides multi-omics evidence (transcriptomic, proteomic, and metabolomic) for the involvement of genes in AD. The platform enables access to data generated by multiple collaborative research programs and was built to support AD target discovery. Agora (https://agora.adknowledgeportal.org/) provides a comparison tool for differential ribonucleic acid (RNA) and protein expression, comprehensive summaries of evidence for individual genes, and lists of genes nominated as potentially promising targets by different research teams.

- **AlzGPS** [310] – is a multi-omics platform built to enable genome-informed target identification and drug repositioning in AD. AlzGPS (https://alzgps.lerner.ccf.org) provides access to various omics datasets, including genomics, transcriptomics, proteomics, and metabolomics, as well as information on FDA approved drugs and AD clinical trials. Omics evidence for single biological entities or predefined datasets (e.g., manually curated list of AD-associated genes) is provided and

interlinked across the platform. Protein-protein interaction (PPI) network visualizations are provided for single genes (1-step neighborhood) or connected components among pre-defined gene sets. The resource includes datasets from human studies and studies in model organisms, partly with small sample sizes (n<20).

- **AlzCode** [311] – is a results explorer that allows users to evaluate the degree of association to known AD genes for individual genes. This is done by providing genomic evidence, including gene expression data and homogeneous biological networks (e.g., co-expression networks and PPIs) taken from public databases. AlzCode (http://www.alzcode.xyz/) also provides a tool to evaluate the interaction of a given set of genes with known AD genes on the bases of different biological networks.

In addition, network-based resources, in which the relationships between biological entities and diseases are modeled as a large heterogeneous network, are becoming increasingly important due to their intuitive interpretation and the abundance of downstream analysis tools. Here, links (edges) between two entities (nodes) indicate either a knowledge-based (e.g., taken from an established knowledge base) or data-driven (e.g., inferred through statistical analysis of quantitative omics data) relationship. For such heterogeneous network resources, the inclusion of high-quality and disease-relevant datasets plays an important role, as it can have a major impact on the networks generated. These resources can be queried through efficient network queries and graph algorithms. Current resources in this category include:

- **HetioNet** [204,312] – is a biomedical knowledge graph that integrates various public data resources covering basic biological, disease, and pharmacological relationships that link entities such as diseases, genes, and drug compounds. This resource covers multiple diseases and is not AD-specific. It further does not include metabolomics data. The database is freely accessible, and pathfinding tools and results from gene prioritization and drug repositioning analysis are available through the web interface (https://het.io/).

- **NeDRex** [313]– is another integrative, network-based resource developed for drug repositioning and disease module identification. This resource integrates various biomedical databases, establishing links between biomedical entities such as diseases, genes, and drug compounds. The platform provides API endpoints to access the integrated data directly and various network algorithms implemented as a Cytoscape plugin (https://nedrex.net/index.html). NeDRex is not AD-specific and does not include metabolomics data.

- **HENA** [188] – is an AD-specific heterogeneous network-based dataset that has been made available through the Network Data Exchange (NDEx) repository [101] (https://doi.org/10.18119/N93G6T). This collection of genomics, proteomics, and clinical phenotype data integrates various interaction datasets, including PPI, GWAS, co-expression, and epistatic interaction networks. All data was projected onto genes, resulting in a gene-gene interaction network connected by different edge types. This dataset does not include metabolomics data and provides no custom analysis tools.

While the presented resources provide valuable tools for the integrative analysis of complex diseases, many have so far a) not leveraged the available multi-omics data to reconstruct data-driven network representations that can be integrated across; b) do not integrate metabolic readouts, or c) do not provide tools to explore several hypotheses in one integrated context network, which poses a significant limitation for multi-factorial diseases like AD.

Material and methods 4

Parts of this Chapter have been used to prepare the following two first author manuscripts (*peer-reviewed journal submission in preparation*):

1. M. A. Wörheide, J. Krumsiek, S. Nataf, K. Nho, A. K. Greenwood, J. C. Wiley, T. Wu, K. Huynh, P. Weinisch, W. Römisch-Margl, N. Lehner, The AMP-AD Consortium, The Alzheimer's Disease Neuroimaging Initiative, The Alzheimer's Disease Metabolomics Consortium, J. Baumbach, P. J. Meikle, A. J. Saykin, P. Murali Doraiswamy, C. van Duijn, K. Suhre, R. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"An online molecular atlas of Alzheimer's disease"**, *submission in preparation*. Preprint available on medRxiv, doi: 10.1101/2021.09.14.21263565.
2. M. A. Wörheide, J. Krumsiek, S. Nataf, K. Nho, A. K. Greenwood, J. C. Wiley, T. Wu, K. Huynh, P. Weinisch, W. Römisch-Margl, N. Lehner, The AMP-AD Consortium, The Alzheimer's Disease Neuroimaging Initiative, The Alzheimer's Disease Metabolomics Consortium, J. Baumbach, P. J. Meikle, A. J. Saykin, P. Murali Doraiswamy, C. van Duijn, K. Suhre, R. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"Utilizing multi-omics context networks to explore molecular hypotheses in Alzheimer's disease"**, *submission in preparation*. Preprint available on medRxiv, doi: 10.1101/2021.09.14.21263565.

Contributions of the lead author for both manuscripts were as follows:

- implementation of the project
- data analysis and resource development (backend and frontend)
- drafting the manuscript
- design and implementation of all figures

Critical revision of the manuscripts and final approval was given by all.

The following chapter provides an overview of Alzheimer's disease (AD) cohorts, datasets, and methods that were used to create the AD Atlas resource (Chapter 5). First, key AD cohorts, initiatives, and hubs for AD data exchange are briefly discussed, followed by a comprehensive list of datasets that were integrated into the resource. Next, an outline of the applied multi-omics data integration pipeline is given, followed by a description of the downstream analysis methods that were used to explore and analyze the resulting correlation-based multi-omics networks. Finally, the underlying technical framework of the AD Atlas front- and backend is described.

4.1 AD data

4.1.1 Cohorts

Data from the following AD cohorts was reanalyzed prior to integration into the AD Atlas. As samples may have varying availability of omic and biomarker profiles, the sample sizes reported refer to the total (maximum) sample size, although some analyses were performed on subsets thereof. An overview of the available omics data modalities for each of these cohorts is given in Table 4.1. A comprehensive list of the integrated datasets and methods used for re-analyzing the data is provided in Section 4.2.

Religious Orders Study/Memory and Aging Project

The Religious Orders Study (ROS) [314] (started in 1994) is a longitudinal, clinical-pathological cohort study that enrolls participants without signs of dementia from religious communities (for example, nuns, priests, and brothers) from across the United States. Participants undergo annual evaluations and have all consented to brain donation after death. The Memory and Aging Project (MAP) [315] (started 1997) is a longitudinal and epidemiological, clinical-pathological cohort. The study design is similar to ROS; enrolled participants are dementia-free at baseline and undergo a detailed annual clinical evaluation and the donation of the brain, spinal cord, and muscle after death. In contrast to ROS, this study recruits a broader range of socioeconomic backgrounds, especially regarding education. Data generated from post-mortem brain tissue samples (dorsolateral prefrontal cortex (DLPFC)) were used to estimate protein partial correlation networks (tandem mass tag mass spectrometry (TMT-MS)-based proteomics data, n=328) and perform genome-wide association studies (GWAS) with brain metabolite levels (genomics and non-targeted metabolomics data, n=459). For the latter, data was also meta-analyzed with results from the Mayo Clinic Brain Bank (Mayo).

Alzheimer's Disease Neuroimaging Initiative

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal, multi-site study of normal aging, Mild Cognitive Impairment (MCI) and early AD. It was funded in 2004 as a private-public partnership through federal (National Institutes on Aging (NIA)), industry, and foundation contributions, to develop neuroimaging measures and biomarkers for clinical trials, and advance the understanding of AD pathophysiology [316]. Participants undergo extensive clinical and cognitive assessments upon entry, including brain imaging (magnetic resonance imaging (MRI) and positron emission tomography (PET) scans) as well as biospecimen collection (urine, blood, cerebrospinal fluid (CSF)). These tests are repeated in specific intervals throughout years of follow-up. The initial study (ADNI-1) has been sequentially extended through additional funding (ADNI-GO) and re-newel of the initial grant (ADNI-2). The following provides a short overview of each ADNI phase and their respective focus:

ADNI-1 The initial five-year study enrolled 800 participants (elderly controls, MCI, AD) during the years 2004-2010. The primary goal was to develop biomarkers that can function as outcome measures in clinical trials.

ADNI-GO Extension of ADNI-1 until 2011 enabled the ADNI-GO phase to recruit 200 additional participants (early MCI) to focus on biomarkers for early stages of the disease.

ADNI-2 During 2011-2016 additional participants (n=550) were recruited including elderly controls, early MCI, late MCI and AD). In addition to assessing biomarkers as outcome measures, the predictive value of these measures for cognitive decline or dementia was added as a subject of interest [317].

Data and results generated by ADNI are shared without embargo through the Laboratory of Neuroimaging (LONI) Image & Data Archive (IDA) [3, 318]. Samples from ADNI-1, -GO, and -2 were used to estimate metabolite partial correlation networks (targeted metabolomics data, n=1,517) and perform GWAS with blood metabolites (targeted metabolomics data, n=1,407) and AD (endo-)phenotypes including CSF biomarkers, cognitive measures and clinical diagnosis (n=1,564).

Mayo Clinic Brain Bank

The Mayo Clinic Brain Bank (Mayo) is a brain bank for neurodegenerative disorders, including AD and Lewy body disease, operated by the Neuropathology and Microscopy Laboratory of the Mayo Clinic. Well-characterized brain samples are stored and distributed to researchers for various types of research studies. Data from post-mortem brain tissue samples (temporal cortex (TCX), n = 159; cerebellum (CBE), n = 177) were used to identify genetic associations with brain metabolites (non-targeted metabolomics data) and perform inverse-weighted meta-analysis with results generated from ROS/MAP.

Baltimore Longitudinal Study of Aging

The Baltimore Longitudinal Study of Aging (BLSA) [319] is one of the largest and longest-running prospective studies of aging in the United States of America [320]. The study is supported by the NIA and was started in 1958 to gain deeper insights into the human aging process. Participants are assessed at regular intervals, measuring physical and cognitive changes. Furthermore, an autopsy program was established in 1986, enabling the study of morphological and neuropathological changes in neurodegenerative diseases [321]. label-free quantitation mass spectrometry (LFQ-MS)-based proteomics data of post-mortem brain tissue samples (DLPFC, n=37) were reanalyzed together with data from Adult Changes in Though (ACT), Banner Sun Health Research Institute (Banner) and Mount Sinai Brain Bank (MSBB) to estimate protein partial correlation networks.

Table 4.1: Multi-omics profiling of selected Alzheimer's disease cohorts.

					phenotypic		
	G	T*	P	M	CSF†	neuropathalogical‡	clinical
ROSMAP	x	DLPFC	DLPFC	DLPFC		x	x
ADNI (1/GO/2)	x			blood	x	x	x
Mayo	x	CBE TCX		CBE TCX		x	x
Banner			DLPFC			x	x
BLSA			DLPFC			x	x
ACT			DLPFC			x	x
MSBB	x	FP IFG PHG STG	DLPFC			x	x

Cohorts – ROS/MAP: Religious Orders Study/Memory and Aging Project, ADNI: Alzheimer's Disease Neuroimaging Initiative, Mayo: Mayo Clinic Brain Bank, Banner: Banner Sun Health Research Institute, BLSA: Baltimore Longitudinal Study of Aging, ACT: Adult Changes in Though, MSBB: Mount Sinai Brain Bank. Brain regions – DLPFC: dorsolateral prefrontal cortex, CBE: cerebellum, TCX: temporal cortex, FP: frontal pole, IFG: inferior frontal gyrus, PHG: parahippocampal gyrus, STG: superior temporal gyrus. Data modalities – G: genomics, T: transcriptomics, P: proteomics, M: metabolomics. *Data not reanalyzed prior to integration. †CSF biomarker measurements. ‡Post-mortem (neuropathological assessment) or ante-mortem (imaging-based).

Mount Sinai Brain Bank

The Mount Sinai/JJ Peters VA Medical Center National Institutes of Health (NIH) Brain and Tissue Repository [322] is a biorepository for donated brain tissue samples, a so-called brain bank. As part of the NIH-funded NeuroBioBank, it facilitates the collection, storage, and distribution of well-characterized post-mortem brain tissue samples. It was established in 1982 with a focus on AD and later expanded to include other brain-associated disorders, including, for example, schizophrenia. For the MSBB study, genomic, transcriptomic, and proteomic data from multiple brain regions are available, as well as demographic and neuropathological data. LFQ-MS-based proteomics data of post-mortem brain tissue samples (DLPFC, n=159) was reanalyzed together with data from ACT, Banner and BLSA to estimate protein partial correlation networks.

Banner Sun Health Research Institute

The Brain and Body Donation Program at Banner is a longitudinal clinical-pathological study established in 1987 of AD, Parkinson's disease, and normal aging [323]. The study collects and stores brain- and whole-body donations of donors from the greater Phoenix metropolitan area. Most participants are enrolled while cognitively normal, although the program also accepts and specifically recruits donors living with neurodegenerative disorders, such as AD and Parkinson's disease. LFQ-MS-based proteomics data of post-mortem brain tissue samples (DLPFC, n=167) was reanalyzed together with data from MSBB, ACT and BLSA to estimate protein partial correlation networks. TMT-MS-based proteomics data from the Banner cohort (DLPFC, n=160) was additionally used in a second analysis to estimate protein partial correlation networks together with data from ROS/MAP.

Adult Changes in Thought

The ACT study is a longitudinal prospective cohort study that enrolls randomly selected members of the Kaiser Permanent Washington health plan. The study was initiated in 1994 to provide estimates of the incidence of dementia and AD, and to examine the influence of sex, education, and apolipoprotein E (*APOE*) status on the onset of dementia [324]. Participants eligible to enroll in the study must be at least 65 years of age and present no signs of dementia. Consent to brain donation after death is not mandatory. Assessment is carried out through interviews and cognitive screening at baseline and then biennially (annually after dementia diagnosis). LFQ-MS-based proteomics data of post-mortem brain tissue samples (DLPFC, n=56) was reanalyzed together with data from MSBB, Banner and BLSA to estimate protein partial correlation networks.

4.1.2 Accelerating Medicines Partnership - Alzheimer's Disease

The Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) is a precompetitive, public-private partnership led by the NIA. It is part of the Accelerating Medicines Partnership Program, which was launched in 2014 with the overall goal of reducing the time and cost of drug development by bringing together government, industry, and nonprofit organizations. Collaborators within the AMP-AD program have generated large-scale quantitative datasets for numerous cohorts (Section 4.1.1) by applying a collection of modern technologies and analyzing this data to identify and validate potential therapeutic targets. For example, the Alzheimer's Disease Metabolomics Consortium (ADMC) has generated broad metabolomics profiles of brain tissue samples of participants of the ROS/MAP and Mayo cohorts. Following the Open Data paradigm, all data, standardized pre-processing and analysis pipelines, and results are shared via a central data infrastructure (see Section 4.1.3). The AMP-AD program has so far undergone two iterations:

AMP-AD 1.0 The first iteration of the program was launched in 2014, consisting of two international and interdisciplinary projects that were focused on the evaluation of tau imaging as a biomarker for clinical trials and disease prognosis ("Biomarkers in Clinical Trials Project") and the discovery of novel therapeutic targets through the integration of large-scale omics data and systems biology approaches ("Target Discovery and Preclinical Validation Project").

AMP-AD 2.0 In the second iteration of the program (started in 2021) the "Target Discovery and Preclinical Validation Project" was expanded, with an emphasis on precision medicine via biomarker discovery, and longitudinal and single-cell datasets.

4.1.3 Data sharing repositories

AD Knowledge Portal

The AD Knowledge portal (https://adknowledgeportal.org/; [325]) is a public data repository that was initially established as part of the AMP-AD program but has now expanded to include additional AD- and aging-related research programs. It was developed to provide a centralized data infrastructure that facilitates the distribution of data, methods, and results generated through NIA-funded research efforts under FAIR (Findable, Accessible, Interoperable and Reusable; [227]) principles. Data files containing primary data, analysis results, and meta-analysis can be directly downloaded via the underlying data storage solution Synapse [326]. Data generated within the AMP-AD program that was downloaded and integrated into the AD Atlas includes, but is not limited to, brain expression quantitative trait loci (eQTL) data [327], brain gene co-expression [328] and brain proteomics data [329, 330], as well as differential analysis results of the brain transcriptome [328] and proteome [329, 330].

NIA Genetics of Alzheimer's Disease Data Storage Site

The NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) (https://www.niagads.org/; [331]) is a data repository that provides harmonized access to data generated by NIA-funded genetic studies. Besides genomic, genetic, and phenotypic data, this platform also provides access to summary statistics of large-scale association studies of Alzheimer's disease and related dementia's (ADRD) that have been published in peer-reviewed journals. AD GWAS were downloaded and integrated into the AD Atlas.

4.2 Integrated datasets

In the following, datasets that are currently integrated into the AD Atlas are listed and additional information regarding their generation and preprocessing is given.

4.2.1 Public databases and population-based studies

Ensembl

Ensembl [332] is a bioinformatics resource that was built to store, annotate and display genome information [333]. The AD Atlas uses the Ensembl database to establish knowledge-based relationships that link genes, transcripts and proteins. Furthermore, Ensembl identifiers (IDs) are used as the primary and unique ID for genes, transcripts and proteins. Ensembl version 97 was accessed in April 2019 using the R package biomaRt [89]. Genes are further annotated with approved human gene nomenclature provided by the HUGO Gene Nomenclature Committee (HGNC) [334]. The complete set of annotations (gene id and gene symbol) was downloaded from ftp://ftp.ebi.ac.uk/pub/databases/genenames/new/tsv/hgnc_complete_set.txt.

SNiPA

The single nucleotide polymorphism (SNP) annotation database SNiPA [148] v3.3. was used to project SNPs to genes using multiple layers of information, including the genomic location, quantitative trait loci (QTL), such as eQTL associations (cis- and trans-) and protein quantitative trait loci (pQTL) associations (cis- and trans-), and gene-associated regulated elements (ENCODE [335], FANTOM5 [336]) [26]. SNP-to-gene mappings determined by SNiPA were downloaded and included in the AD Atlas. Furthermore, this mapping was used to determine the gene-specific significance threshold described in Section 4.3.2.

Tissue-specific gene regulation

Genotype-Tissue Expression (GTEx) Project (2020)

Information on genetic loci that affect the expression of protein-coding genes was taken from the Genotype-Tissue Expression (GTEx) project [143]. The version 8 (v8) data release examines n=15,201 ribonucleic acid (RNA)-sequencing samples from 49 tissues of 838 postmortem donors with genotype data from whole-genome sequencing (WGS). Donors were primarily (85.3%) European Americans. Significant variant-gene associations (cis- and trans-) based on permutations were downloaded from the GTEx Portal for each tissue (GTEx_Analysis_v8_eQTL.tar and GTEx_Analysis_v8_trans_eGenes_fdr05.txt). GTEx variant IDs were mapped to reference SNP cluster IDs (rsIDs) prior to integration using the provided lookup table for genotyped variants (GTEx_Analysis_2017-06-05_v8_WholeGenomeSeq_838Indiv_Analysis_Freeze.lookup_table.txt.gz). Sample sizes for brain tissues with RNA-Seq and genotype data available are given in Table 4.2 [143].

Table 4.2: Sample sizes for brain region-specific eQTLs taken from GTEx [143].

Tissue Type	Samples with RNA-seq and genotype
Brain - Amygdala	129
Brain - Anterior cingulate cortex (BA24)	147
Brain - Caudate (basal ganglia)	194
Brain - Cerebellar Hemisphere	175
Brain - Cerebellum	209
Brain - Cortex	205
Brain - Frontal Cortex (BA9)	175
Brain - Hippocampus	165
Brain - Hypothalamus	170
Brain - Nucleus accumbens (basal ganglia)	202
Brain - Putamen (basal ganglia)	170
Brain - Spinal cord (cervical c-1)	126
Brain - Substantia nigra	114

Sieberts et al. (2020) – Brain cis-eQTL meta-analysis

Sieberts et al. [327] conducted a large-scale analysis of cortical cis-eQTL using n=1433 samples from four cohorts from the AMP-AD Consortium and the CommonMind Consortium. eQTLs were generated separately for each cohort/tissue, adjusting for diagnosis and principal components of ancestry. Then meta-analysis was performed via a fixed-effect model. This included following brain regions: DLPFC from ROS/MAP, MSSM-Penn-Pitt, and HBCC, and TCX from Mayo. Results of the cortical meta-analysis were downloaded from Synapse. A significance cutoff of FDR(P_{meta}) ≤ 0.05 was applied and SNP rsIDs and Ensembl IDs (for genes) were used for integration.

Genetic associations with metabolic traits (mGWAS)

Summary statistics from GWAS with metabolic traits were downloaded via SNiPA [148]. Results filtered to a default p-value cutoff of $P \leq 1 \times 10^{-4}$ were available, if not stated otherwise. Where possible, IDs for genes and metabolites were mapped to Ensembl or metabolomics platform-specific IDs, respectively. The included studies cover metabolic traits measured in serum, plasma, and urine samples using both targeted and non-targeted metabolomics approaches.

Suhre et al. (2011) - Human metabolic individuality

Suhre et al. [29] profiled fasting serum samples from participants of the KORA F4 study (n=1,768) and the TwinsUK study (n=1,052) using the non-targeted metabolomics platform Metabolon. After quality control 276 (KORA) and 258 (TwinsUK) metabolites were used for further analysis, respectively. Linear models were fitted for each cohort separately

to log-transformed metabolic traits and adjusted for age, gender, and family structure. Association results for both of these analyses were downloaded and integrated into the AD Atlas. SNP rsIDs and Metabolon-specific compound identifiers (COMP IDs) for metabolites were used for integration.

Shin et al. (2014) - Atlas of genetic influences on blood metabolites

Shin et al. [25] investigated how genetic variation influences metabolism and complex traits using data from n=7,824 individuals from two population-based studies (Kora and TwinsUK). In total, 486 metabolite concentrations profiled in either plasma or serum were present in both datasets after quality control. The authors tested for associations between each SNP and metabolite concentration using linear regression models adjusted for age and sex. Batch effect was only added to the model for the TwinsUK analysis. Both cohorts were analyzed separately, using the software QUICKTEST in KORA and Merlin in TwinsUK. The resulting cohort-level summary statistics were combined using inverse variance meta-analysis based on effect size estimates and standard errors, adjusting for genomic control. Association data for all 486 metabolites were downloaded and SNP rsIDs and Metabolon-specific COMP IDs for metabolites were used for integration.

Raffler et al. (2015) - Loci of urinary human metabolic individuality

Raffler et al. [26] performed a GWAS using metabolically characterized urine samples (targeted metabolomics) and genotype data available for n=3,861 study participants from the longitudinal population study SHIP-0. The software PLINK (v1.07) was used to fit age- and sex-corrected linear regression models and a Bonferroni-adjusted significance threshold was applied to correct for multiple testing. The association results for 55 targeted metabolic traits in urine were included in the AD Atlas. SNP rsIDs and biochemical names for metabolites were used for integration.

Draisma et al. (2015) - Genetic variants contributing to variation in blood metabolite levels

Draisma et al. [24] performed a meta-analysis of genome-wide association analysis using blood serum samples from n=7,478 individuals across seven cohorts and five countries (the Netherlands, Germany, Australia, Estonia, and the United Kingdom). Genotype data and targeted metabolomic measurements (129 metabolites), performed using the Biocrates platform, were first analyzed assuming a linear model of association, adjusting for age, sex, relatedness, and study-specific covariates as necessary. In the next step, these cohort-level summary statistics were pooled in an inverse variance-weighted, fixed-effects meta-analysis using the software METAL. The association results were downloaded and SNP rsIDs and biochemical names metabolites were used for integration.

Long et al. (2017) - Common-to-rare variants associated with human blood metabolites

Long et al. [337] conducted a whole-genome sequencing study of common, low-frequency, and rare variants in n=1,960 participants of the TwinsUK study. Furthermore, the authors profiled serum samples from these participants collected at three clinical visits using the non-targeted metabolomics platform Metabolon. Focusing on 644 metabolites that showed consistent levels across these data collections and measurable heritability, a linear mixed model was applied to test for associations between genetic variants and metabolite levels while accounting for family structure. As a quantitative trait the $\log_{10}$-transformed mean of the median-normalized values from three visits was used. Sex and mean age at serum collection were included as covariates. Summary statistics for each of the 644 metabolites were downloaded ($P \leq 1 \times 10^{-5}$) and included in the AD Atlas. rsIDs and biochemical names for SNPs and metabolites were used for integration, respectively.

Partial correlation networks

Krumsiek et al. (2012) - A systems approach to metabolite identification

Krumsiek et al. [168] combined GWAS and Gaussian Graphical Models (GGM) to identify previously unknown measured metabolites. GGMs, which are based on partial correlation coefficients, were applied to a dataset of 517 metabolic traits and genotype information on 655,658 genetic variants measured in n=1,768 fasting serum samples from the German population cohort KORA. Confounding through age, gender, and SNP effects were removed by including these variables in the constructed linear models. Partial correlations between metabolite pairs that are significantly different from zero at $\alpha = 0.05$ after Bonferroni correction ($P \leq 7.96 \times 10^{-7}$) and $abs(cor) \geq 0.1603$, were downloaded from the supporting information of the paper and included in the AD Atlas using metabolite biochemical names for the integration.

Suhre et al. (2017) - GWAS with the human blood plasma proteome

Suhre et al. [146] performed a large-scale proteomics-based genetic association study in the German cohort KORA, identifying associations between genetic variants and protein levels, measured by the aptamer-based proteomics platform SOMAscan. In the scope of the study, a GGM was computed for n=997 blood plasma samples using unscaled data for 1,124 proteins and correcting for age, gender, and body mass index (BMI). Partial correlation edges between proteins were included in the network after applying a significance threshold of $\alpha = 0.05$ after Bonferroni correction for all possible edges in the model ($P \leq 7.9 \times 10^{-8}$). Supplementary dataset 2 (annotation of the SOMAmer probes) and dataset 3 (GGM edges) were downloaded. Using the annotations supplied, SOMAmer probe identifiers were mapped to Uniprot IDs (in some cases multiple) for subsequent integration into the AD Atlas.

Table 4.3: Data sources and references. A comprehensive list of data that was integrated into the AD Atlas with the respective significance threshold.

Data	Edge Type	Significance cutoff	Cohort	Reference
GGM	COABUNDACE	0.05/all possible edges	KORA	[146]
		q-value < 0.05 and p-value(cor) significant	BLSA ACT MSSB Banner	[329]
		q-value < 0.05 and p-value(cor) significant	ROSMAP Banner	[330]
	COEXPRESSION	see publication	ROSMAP MSBB Mayo	[328]
	DEG	$P_{FDR} \leq 0.05$	ROSMAP MSBB Mayo	[328]
	DEP	$P_{Tukeys} \leq 0.05$	BLSA ACT MSSB Banner	[329]
		$P_{Holm} \leq 0.05$	ROSMAP Banner	[330]
GGM	PARTIAL CORRELATION	$P \leq 7.96e^{-7}$ and abs(cor) ≤ 0.1603	KORA	[168]
		$P_{Bonf} \leq 0.05$	ROSMAP	[263]
		0.05/all possible edges	ADNI	This work
		0.05/all possible edges	ADNI	This work
eQTL	COREGULATION	q-value ≤ 0.05	GTEx	[143]
		$P_{FDR} \leq 0.05$	AMP-AD CMC	[327]
mQTL	GENETIC ASSOCIATION	$P \leq 0.0001$ #	KORA	[29]
		$P \leq 0.0001$ #	TwinsUK	[29]
		$P \leq 0.0001$ #	KORA TwinsUK	[25]
		$P \leq 0.0001$ #	SHIP-0	[26]
		$P \leq 0.0001$ #	Meta-analysis*	[24]
		$P \leq 0.00001$ #	TwinsUK	[337]
		$P \leq 0.05$ #	ROSMAP	This work

		$P \leq 0.05$ #	ROSMAP Mayo	This work
		$P \leq 0.05$ #	ADNI	This work
MWAS	METABOLIC ASSOCIATION	$P \leq 0.05/15$	ADNI	[280, 281]
		$P \leq 0.05/55$	ADNI	[245]
		$P_{FDR} \leq 0.05$	ROSMAP	[263]
traitQTL	GENETIC ASSOCIATION	$P \leq 0.05$ #	ADNI	This work
		$P \leq 0.05$ #	Meta-analysis**	[278]
		$P \leq 0.05$ #	Meta-analysis*	[338]
		$P \leq 0.05$ #	ADNI Knight ADRC	[339]
		$P \leq 0.05$ #	Meta-analysis*	[340]
		$P \leq 0.05$ #	Meta-analysis**	[341]
		$P \leq 0.05$ #	IGAP UK Biobank	[275]
		$P \leq 0.05$ #	IGAP	[275]
		$P \leq 0.05$ #	Meta-analysis**	[272]
		$P \leq 0.05$ #	PGC-Alz IGAP ADSP UK Biobank	[274]
		$P \leq 0.05$ #	Meta-analysis*	[342]
		$P \leq 0.05$ #	Meta-analysis*	[273]

*meta-analysis across more than four different cohorts; **meta-analysis across International Genomics of Alzheimer's Project (IGAP); #results up to reported threshold are integrated in resource and genome-wide or gene-wide significance is applied to generate context-specific networks

4.2.2 Alzheimer's disease-related associations

Genetic associations with Alzheimer's (endo-)phenotypes

Summary statistics for the following large-scale genome-wide association studies on CSF biomarkers, including amyloid-β, tau and clusterin, and neuropathological burden, as well as case-control meta-analyses, were integrated into the AD Atlas. Datasets were downloaded from the NIAGADS (https://www.niagads.org/) if not stated otherwise. SNP rsIDs were used as unique identifiers and AD (endo-)phenotypes were manually harmonized. Phenotype descriptions are given in Table 8.3 in the Appendix (Chapter 8).

Lambert et al. (2013) - Two-stage meta-analysis of GWAS with late-onset AD

Lambert et al. [278] performed a meta-analysis across four published AD case-control studies from the IGAP consortium, including the Alzheimer's Disease Genetics Consortium (ADGC), Genetic and Environmental Risk in Alzheimer's Disease (GERAD), European Alzheimer's Disease Initiative (EADI), and Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE). In total, n=54,162 samples were included in the meta-analysis, which was performed using a fixed-effects inverse variance-weighted method. Before the combination of the summary statistics the study-specific genomic inflation factors were estimated and their square roots were used to scale the standard errors of the beta coefficient. Stage 1 summary statistics were integrated into the AD Atlas.

Beecham et al. (2014) - GWAS meta-analysis of AD neuropathology

Beecham et al. [338] performed a GWAS using harmonized neuropathological data and genotyping data from n=4,914 brain autopsies. Samples were contributed by the NIA Alzheimer's Disease Centers (ADCs) and ADGC-collaborating studies. Genome-wide association analysis was performed for 14 traits in total, including neuropathological AD features and other brain pathologies. The analysis was performed for each cohort separately. For binary traits, logistic regression was performed and for ordinal traits, polytomous logistic regression. The first three principal components were included as covariates to account for population structure. Inverse-weighted meta-analysis was performed afterward using METAL, accounting for small sample sizes and incomplete phenotyping data by only regarding specific sets for analysis (further described in [338]).

Deming et al. (2016) - GWAS with cerebrospinal fluid clusterin

Deming et al. [339] used cerebrospinal fluid clusterin (CLU) levels as an endophenotype for a GWAS using data from n=673 individuals from the Charles F. and Joanne Knight Alzheimer's Disease Research Center (Knight ADRC) (n=400) and the ADNI (n=273). ADNI and Knight ADRC datasets were combined and the log-transformed and standardized values tested for normality. To test for the association of CSF clusterin levels, statistical analysis was performed using an additive model in PLINK, and study, age, gender, and the first two principal components were used as covariates.

Deming et al. (2017) - GWAS with Alzheimer's endophenotypes

Deming et al. [340] performed a genome-wide association analysis of three endophenotypes (CSF levels of amyloid beta - Aβ_{42}, tau and phosphorylated tau - ptau$_{181}$) using data collected from n=3,146 participants across nine studies, including the Knight ADRC, ADNI, Predictors of Cognitive Decline Among Normal Individuals (BIOCARD) and the Mayo clinic. Raw protein levels were log-transformed and normalized within each study before combination for association testing using an additive linear regression model. Study, age,

sex, and the first two principal components were tested for confounding using a step-wise regression analysis and included as covariates for each protein where applicable

Huang et al. (2017) - GWAS with AD age of onset

Huang et al. [341] conducted a genome-wide survival association study to identify genetic loci associated with AD age at onset. Genotyped samples (n=40,255) from IGAP, including the ADGC, GERAD, EADI, and CHARGE, were used to perform genome-wide Cox proportional hazards regression using an additive model. Sex, site, and the first three (four for EADI) principal components from EIGENSTRAT were included as covariates in all models. This analysis was performed for each dataset separately and then combined using inverse-variance meta-analysis using METAL.

Marioni et al. (2018) - GWAS on family history of Alzheimer's disease

Marioni et al. [275] used samples from the UK Biobank (UKBB) cohort to conduct a GWAS using an AD-proxy phenotype. Analysis was conducted separately for maternal and paternal AD. GWAS was conducted using an additive model and as outcome, the residuals of a linear regression model of maternal/paternal AD status on age of parent (at death or self-report), assessment center, genotype batch, array, and genetic principal components were used. The results were subsequently combined in two meta-analyses, performed using a standard error-weighted meta-analysis in METAL; 1) meta-analysis of UKBB maternal and paternal analysis (n= 314,278) and 2) meta-analysis of UKBB and published stage II summary statistics from IGAP [278] (total n = 388,324). Summary statistics of both meta-analyses are included in the AD Atlas.

Kunkle et al. (2019) - Meta-analysis of diagnosed Alzheimer's disease

Kunkle et al. [272] conducted a genetic meta-analysis across consortia of IGAP, which includes the ADGC, GERAD, EADI, and CHARGE. In total, the analysis included 35,274 AD cases and 59,163 controls (total n=94,437). In discovery stage 1 an additive genotype model was used to test for associations between case-control status and genotype for each dataset. Models were adjusted for age (age at onset or age at last exam), sex and principal components. Results across cohorts were combined using inverse-variance meta-analysis as implemented in METAL. In stage 2 replication analysis was carried out using a custom genotyping chip [278] and stages 3A and 3B provided further replication for variants in regions not well captured by the chip. Summary statistics for stages 1, 2 (combined stage 1 and stage 2 p-values), 3A, and 3B were downloaded and merged, such that p-values measured in later stages (2, 3A, and 3B) were reported (if available) for each SNP.

Jansen et al. (2019) - Meta-analysis of clinically diagnosed and proxy AD

Jansen et al. [274] performed a large-scale genome-wide association analysis using both clinically diagnosed AD and AD-by-proxy cases (total n=455,258). This included n=79,145 samples with clinically diagnosed AD case-control status across three consortia; Alzheimer's disease working group of the Psychiatric Genomics Consortium (PGC-ALZ), IGAP, and the Alzheimer's Disease Sequencing Project (ADSP), as well as n=376,113 samples with a weighted AD-by-proxy phenotype from the UKBB. Here, parental AD status from self-report questionnaires was used to generate a score where the number of affected parents was included and unaffected parents were weighed by their age or age at death. Association analyses were performed for each cohort using linear (UKBB - adjusted for ancestry principal components, age, sex, genotyping array, assessment center) and logistic regression (ADSP - adjusted for sex, batch, ancestry principal components; PGC-ALZ - adjusted for sex, batch, ancestry principal components, and age). For IGAP, summary data was used. All cohorts were meta-analyzed using a multivariate genome-wide meta-analysis that takes into account the partial overlap between cohorts by defining a custom per SNP test statistic. Summary statistics from this analysis (phase 3) were downloaded and integrated.

Wightman et al. (2021) - Meta-analysis of clinically diagnosed and proxy AD

Wightman et al. [342] extended the analysis of Jansen et al. [274] to include samples from additional cohorts, including Finngen and GR@CE. A comprehensive list of cohorts can be found in [342]. Analysis and meta-analysis were conducted as described above for Jansen et al. using a total of n=762,917 samples. Summary statistics, excluding the 23andMe data, were downloaded and integrated.

Bellenguez et al. (2022) - Meta-analysis of clinically diagnosed and proxy AD

Bellenguez et al. [273] meta-analyzed a large dataset of clinically diagnosed case-control samples from cohorts across the European Alzheimer & Dementia Biobank (EADB) consortium and proxy-AD/Dementia cases from the UKBB. In total, n=487,511 samples were included; 39,106 clinically diagnosed AD cases, 46,828 proxy-AD/Dementia cases, and 401,577 controls. Proxy-AD/Dementia status was determined using self-report questionnaires and participants were included if at least one biological relative (parent or sibling) was reported to have dementia. Association analysis was performed in each dataset separately using logistic regression and an additive genetic mode. For the UKBB dataset, a logistic mixed model was used. All analyses were adjusted for principal components and genotyping center where necessary. To combine results across studies, METAL was used to perform an inverse-variance meta-analysis. Summary statistics of this analysis (stage I) were downloaded and integrated.

Brain co-expression networks

Wan et al. (2020) - Meta-analysis of the human brain transcriptome

Wan et al. [328] conducted a consensus gene co-expression analysis using RNA-seq data to evaluate and robustly identify AD-related molecular signatures. The authors used data from three postmortem brain studies [322, 343, 344] collected within the AMP-AD consortium. Covariate adjustment was performed for each cohort separately, including diagnosis, sex, identified biological and technical covariates, and donor information (as random effect). Multiple gene co-expression network inference methods to n=2,114 samples. Subsequently, the resulting networks were merged into a single meta-network using an ensemble network inference algorithm. The samples were collected from seven distinct brain regions, resulting in seven tissue-specific meta-co-expression networks - DLPFC, TCX, CBE, inferior frontal gyrus (IFG), superior temporal gyrus (STG), frontal pole (FP), and parahippocampal gyrus (PHG). This data is publicly available via Agora, the results explorer of the AMP-AD Knowledge Portal, and was downloaded from Synapse.

Brain partial correlation networks

Batra et al. (2022) - The landscape of metabolic brain alterations in AD

Batra et al. [263] investigated metabolic changes in the brain using n=500 postmortem DLPFC brain tissue samples from the ROS/MAP cohorts. The samples included n=352 females and n=148 males. In total, out of 500 individuals, n=220 were diagnosed with AD, n=119 MCI and n=153 without cognitive impairment, and eight with other forms of dementia. Non-targeted metabolomics profiling was performed using the Metabolon platform and after quality control 667 metabolites were used to compute a partial correlation-based GGM. P-values of the resulting partial correlations were corrected using the Bonferroni method and partial correlations with $P_{adj} \leq 0.05$ were included in the network. Metabolon-specific COMP IDs were used as metabolite identifiers for integration into the AD Atlas.

Differential analysis in AD cohorts

Wan et al. (2020) - Meta-analysis of the human brain transcriptome

As described above, Wan et al. [328] used RNA-seq data collected across three postmortem brain tissue studies covering seven different brain regions. Differential expression analysis and meta-differential expression analysis across brain regions were performed for n=778 samples using weighted fixed- and mixed-effect linear models. Covariate adjustment was performed for each cohort separately and included diagnosis, sex, identified biological and technical covariates, and donor information (as random effect). Differentially expressed genes were determined as those with an adjusted $P_{FDR} \leq 0.05$. Both brain region-specific and meta-analysis differential expression results were downloaded from Synapse. Four different models were included in the AD Atlas; AD Diagnosis (males and females): changes in

gene expression between AD case and controls; AD Diagnosis x age of death (AOD) (males and females): changes in gene expression between AD cases and controls and whether AOD has an impact; AD Diagnosis x Sex (females only): changes in gene expression between female AD cases and controls; AD Diagnosis x Sex (males only): changes in gene expression between male AD cases and controls.

Johnson et al. (2020) - Consensus proteomic analysis of AD brain

Johnson et al. [329] performed a large-scale consensus proteomics analysis of AD brain across the AMP-AD consortium, including BLSA, Banner, MSBB and ACT. DLPFC tissue samples from a total of 453 control, asymptomatic AD and AD brains were analyzed using LFQ-MS proteomics. After quality control, n=419 samples and 3334 proteins were used for further analysis. Differentially abundant proteins were identified using one-way ANOVA followed by Tukey's comparison post hoc test across control, asymptomatic AD and AD brain tissue samples. The results of this analysis (Paper Supplementary Table 2A) were downloaded. Proteins were categorized as unchanged (Tukey's $P > 0.05$), up-regulated and down-regulated in AD (Tukey's $P \leq 0.05$ and $log_2(FC) > or < 0$, respectively) for the AD Atlas. Uniprot IDs were used for integration.

Johnson et al. (2022) - Large-scale deep multi-layer analysis of AD brain

Johnson et al. [330] performed a large-scale consensus proteomics analysis of AD brain obtained from the autopsy collections of Banner and ROS/MAP. DLPFC and BA9 tissue samples from a total of 516 control, asymptomatic AD and AD brains were analyzed using a TMT-MS proteomics approach. After quality control, n=488 samples and 8619 proteins were used for further analysis. Differentially expressed proteins were identified using one-way ANOVA followed by Holm post hoc correction of all pairwise comparisons. The results of this analysis (Paper Supplementary Table 2A) were downloaded. Proteins were categorized as unchanged (Holm's $P > 0.05$), up-regulated and down-regulated in AD (Holm's $P \leq 0.05$ and $log_2(FC) > or < 0$, respectively) for the AD Atlas. Uniprot IDs were used for integration.

Metabolic associations with Alzheimer's (endo-)phenotypes

Metabolite associations with Alzheimer's (endo-)phenotypes that were integrated into the AD Atlas were identified through metabolome-wide association studies (MWAS) using samples from ADNI phases 1, GO, and 2.

MahmoudianDehkordi et al. (2019) and Nho et al. (2019) - Bile acid MWAS with AD biomarkers

Using n=1555 baseline serum samples of fasting participants of ADNI, including cognitively normal individuals, individuals with early or late MCI, as well as AD cases, tar-

geted metabolomics profiling of bile acids was performed using the Biocrates Life Sciences Bile Acids Kit (BIOCRATES Life Science AG, Insbruck, Austria). Bile acid levels were adjusted for medication effects. Linear regression models were used to analyze the association of the levels of 15 bile acids with 19 AD-related traits, including CSF and imaging biomarkers [280, 281]. Age, sex, study phase, BMI and *APOE**ε4 status were included as covariates. Significant associations that were integrated into the AD Atlas were determined by applying a Bonferroni significance threshold of $P \leq 3.33 \times 10^{-3}(0.05/15)$.

Arnold et al. (2020) - MWAS with AD biomarkers

Using n=1517 baseline serum samples of fasting participants pooled from ADNI phases 1, GO, and 2, Arnold et al. [245] analyzed the association of 19 AD-related traits, including CSF and imaging biomarkers, with the levels of 140 metabolites using standard linear and logistic regression. Metabolite levels were adjusted for significant medication effects using step-wise backward selection. Regression models were adjusted for age, sex, ADNI study phase, and the number of copies of *APOE**ε4 and could also include BMI and education (selected by backward selection). The significant associations that were integrated into the AD Atlas were determined by applying a Bonferroni significance threshold of $P \leq 9.09 \times 10^{-4}(0.05/55)$, as the number of independent metabolic features was determined to be 55.

Batra et al. (2022) - The landscape of metabolic brain alterations in AD

As described above, Batra et al. [263] investigated metabolic changes in the brain using n=500 postmortem DLPFC brain tissue samples from the ROS/MAP cohorts. Non-targeted metabolomics profiling was performed using the Metabolon platform and after quality control 667 metabolites were tested for association with eight AD-related traits: clinical diagnosis at the time of death, level of cognition proximate to death, cognitive decline during lifetime, amyloid-β load, tau tangle load, global burden of AD pathology, NIA-Reagan score and neuropathology diagnosis. Generalized linear models with traits as response variable and appropriate link functions were used to test for association between metabolite concentrations and traits. Age, sex, BMI, post mortem interval (PMI), number of years of education and the number of *APOE**ε4 alleles were included as covariates in the models. P-values were corrected using the Benjamini-Hochberg method to account for multiple testing. Summary statistics for significant associations ($P_{adj} \leq 0.05$) were downloaded and integrated into the AD Atlas using Metabolon specific COMP IDs as metabolite identifiers.

4.2.3 Additional analysis and reanalysis of data

Partial correlation networks

Johnson et al. (2020) - Consensus proteomic analysis of AD brain

As described above (see Section 4.2.2), Johnson et al. [329], performed a meta-analysis of brain tissue across four studies of the AMP-AD consortium. Minimally regressed (batch- and site-corrected) data were downloaded from Synapse and a GGM was estimated to construct a partial correlation network as described in [19], correcting for age, sex, and PMI. Uniprot IDs were used to integrate the data.

Johnson et al. (2022) - Large-scale deep multi-layer analysis of AD brain

As described above (see Section 4.2.2), Johnson et al. [330], performed a meta-analysis of brain tissue across two studies of the AMP-AD consortium. Minimally regressed (batch- and site-corrected) data were downloaded from Synapse and a GGM was estimated to construct a partial correlation network as described in [19], correcting for age, sex, and PMI. Uniprot IDs were used to integrate the data

Alzheimer's Disease Neuroimaging Initiative (ADNI)

Two metabolite partial correlation networks were estimated for metabolites measured by the targeted AbsoluteIDQ-p180 and Bile Acids Kit (BIOCRATES Life Science AG, Innsbruck, Austria) using samples from ADNI [345] phases 1, GO, and 2. Detailed data processing is described in [245] (p180) and [281] (bile acids). In summary, the AbsoluteIDQ-p180 dataset included n=1517 baseline fasting serum samples and 139 metabolites after quality control and the bile acid dataset included a total of 15 bile acids for n=1464 baseline fasting serum samples after quality control. Metabolites were adjusted for medications and dietary supplements. GGMs were estimated for each dataset separately as described in [19], applying a significance threshold of 0.05 after Bonferroni correction for all possible edges in the respective models. Age, sex, number of *APOE**ε4 alleles, and education were included as covariates. Biocrates-given metabolite names were used for integration.

Metabolomics data are available via the AD Knowledge Portal (https://adknowledgeportal.org). AbsoluteIDQ® p180 kit (Biocrates Life Sciences AG, Innsbruck, Austria):

- https://doi.org/10.7303/syn5592519(ADNI-1)
- https://doi.org/10.7303/syn9705278(ADNI-GO/-2)

Bile Acids Kit (Biocrates Life Sciences AG, Innsbruck, Austria):

- https://doi.org/10.7303/syn12036817.1(ADNI-1)
- https://doi.org/10.7303/syn9779093.1(ADNI-GO/2)

Genetic associations with blood metabolites and Alzheimer's (endo-)phenotypes

Alzheimer's Disease Neuroimaging Initiative (ADNI)

Genome-wide genotyping data of ADNI-1/GO/2 participants were collected using the Illumina Human 610-Quad, HumanOmni Express, and HumanOmni 2.5M BeadChips. Before imputation, standard quality control (QC) procedures of GWAS data for genetic markers and subjects were performed (variant call rate < 95%, Hardy-Weinberg-Equilibrium test $P < 1 \times 10^{-6}$, and minor allele frequency (MAF) < 1%, participant call rate < 95%, sex check and identity check for related relatives). Then, non-Hispanic Caucasian participants were selected using HapMap 3 genotype data and MDS analysis. Genotype imputation was performed for each genotyping platform separately using the Haplotype Reference Consortium (HRC) reference Panel r1.1 and merged afterward, resulting in data on n=1,576 individuals and 20,779,509 variants. Using this dataset, we ran GWAS analyses for each outcome (A-T-N-C measures [251], clinical diagnosis, and metabolite levels) that included outcome-specific sets of covariates, including age, sex, study phase, education, and *APOE**ε4 status. All models testing for association with metabolite levels were corrected for clinical diagnosis.

Genetic associations with brain metabolites

We analyzed brain tissue samples of participants of the ROS/MAP cohorts and the Mayo using the non-targeted Metabolon Discovery HD4 platform. ROS/MAP samples were taken from the DLPFC (n = 459), in Mayo samples were taken partly from the TCX (n = 159) and the CBE (n = 177). Metabolic profiles after thorough QC as described in Batra et al. [263] were available for 667 metabolites in ROS/MAP and 658 in Mayo, with an overlap of 576 metabolites available in both datasets. Imputed genotypes were available for all samples. Genotype QC included filters for MAF ≥ 0.05, individual genotyping rate $\geq 95\%$, and genotype call rate $\geq 95\%$, as well as a test $P \geq 1 \times 10^{-5}$ for Hardy-Weinberg-Equilibrium. These filters yielded a total of 4.386 million genotypes, 5.434 million genotypes, and 5.415 million genotypes in ROS/MAP (DLPFC), Mayo (TCX), and Mayo (CBE), respectively. The overlap in genotypes was 4.062 million across datasets. Using z-scored metabolite levels (centered to zero mean and unit variance), we first ran independent genome-wide association studies with metabolite traits (mGWAS) analyses in each brain region for all available metabolites, where the ROS/MAP dataset, being the largest dataset available, was considered to serve as the phase 1 discovery sample. Linear regression was performed for 667 metabolites adjusting for age at death, sex, PMI, and neuropathology-based diagnosis using PLINK2. Replication analyses in the two brain regions available in Mayo TCX, n = 159; CBE, n = 177) were performed analogously for 658 metabolites. Mayo datasets were used as phase 2 replication sets. We afterward conducted an inverse-weighted meta-analysis across all three studies using random-effects models to account for between-study variance caused by differences in cohort recruitment, sample collection, and brain region

specificity of metabolic readouts. We included the summary statistics of the meta-analysis for the 576 overlapping metabolites. For 91 metabolites only measured in the ROS/MAP cohort we included discovery phase p-values. Metabolon-given COMP IDs were used as metabolite identifiers for integration.

4.3 Multi-omics data integration

4.3.1 General integration strategy and data storage

The AD Atlas was built using a step-wise data integration approach in which different omics datasets, such as transcriptomics, proteomics, and metabolomics, are analyzed separately or in specific combinations before integration (shown schematically in Figure 4.1). In contrast to synchronous integration, where all data is used in one analysis step, step-wise approaches allow the integration of data across various sources and do not require the data to come from the same set of individuals/samples [10]. An extended QTL-based integration strategy (Section 2.3.2) paired with a composite network approach (Section 2.3.3) was used. Briefly, relationships between biological entities (e.g., genetic variants, genes, metabolites) are either taken from public knowledge databases or inferred through statistical analysis (e.g., GWAS or correlation-based analysis). These individual networks are subsequently merged into a large heterogeneous network by using overlapping entities

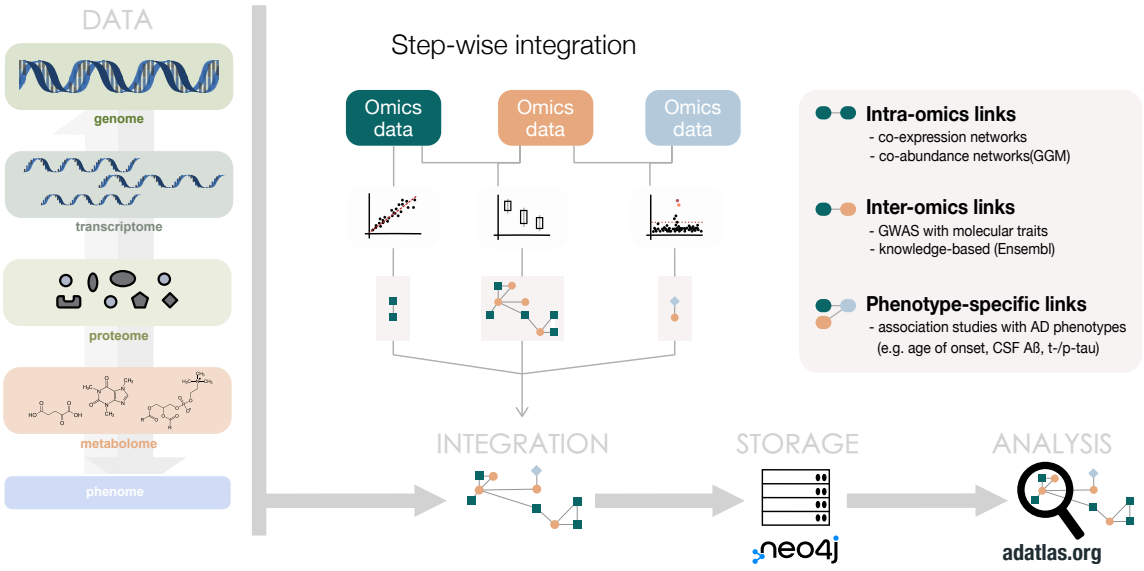


Figure 4.1: Multi-omics data integration strategy. The AD Atlas is a multi-omics resource that enables the integration and analysis of heterogeneous omics datasets in the context of Alzheimer's disease. Multi-omics integration is carried out using a step-wise integration approach. Here, statistical analysis is used to infer biological relationships between (inter-omics) and within (intra-omics) omics layers from omics data collected in large-scale population-based studies. Links to AD (endo-)phenotypes from large-scale case-control or biomarker studies enable multi-omics exploration in the context of AD. The inferred relationships are stored in comprehensive network structures using the graph-based database system Neo4j. Exploration of the data is facilitated through a web interface available at www.adatlas.org.

to connect multiple layers of omics data. For example, links between metabolites and genes are established via overlapping QTLs. The resulting heterogeneous network consists of multiple node types (biological entities) connected by different types of edges (inferred associations between entities). In general, three different types of inferred relationships can be distinguished: inter-omics, intra-omics, and phenotype-specific links (as seen in 4.1):

- **Intra-omics links** establish relationships within one omics layer using the intrinsic correlation structure of omics datasets. In the AD Atlas, we have included brain region-specific gene co-expression networks (gene-gene links), and partial correlation-based protein and metabolite co-abundance networks (protein-protein and metabolite-metabolite links, respectively).
- **Inter-omics links** establish links between omics layers by using knowledge-based gene-transcript-protein links from Ensembl [332] in addition to overlapping genetic associations from GWAS with molecular traits. The AD Atlas has further incorporated tissue-specific eQTL information (taken from GTEx [143], SNiPA [148] and Sieberts et al. [327]), as well as metabolite quantitative trait loci (mQTL) data from large population-based and brain-based studies (listed in Table 4.3).
- **Phenotype-specific links** enable the identification of entities that are associated with AD-specific phenotypes, also referred to as traits, including CSF-biomarkers and neuroimaging measures. They are inferred from large-scale case-control GWAS and MWAS, as listed in Supplementary Table 8.3. This layer of information enables the identification of entities within the network that are associated with AD pathomechanisms.

To enable efficient data storage and analysis, we utilized the native graph database management system Neo4j (https://neo4j.com/). Here, data is stored directly as vertices (nodes) and edges (relationships), allowing data storage that closely aligns with the domain representation [7]. Furthermore, index-free adjacency, i.e., nodes physically pointing to each other, substantially enhances the performance of traversal-type queries (joins) compared to traditional relational databases [346]. A schematic visualization of Neo4j's property graph model is given in Figure 4.2 A.

4.3.2 Data harmonization and integration pipeline

The construction of the AD Atlas can be divided into two main phases. In the first phase, the results of available (multi-)omics studies were collected, preprocessed, integrated, and stored in a highly detailed network structure (Figure 4.2 B). In the second phase, this complex collection of data was summarized and extracted into a simplified data view (Figure 4.2 C) to reduce complexity, increase the interpretability of results and enable efficient data access.

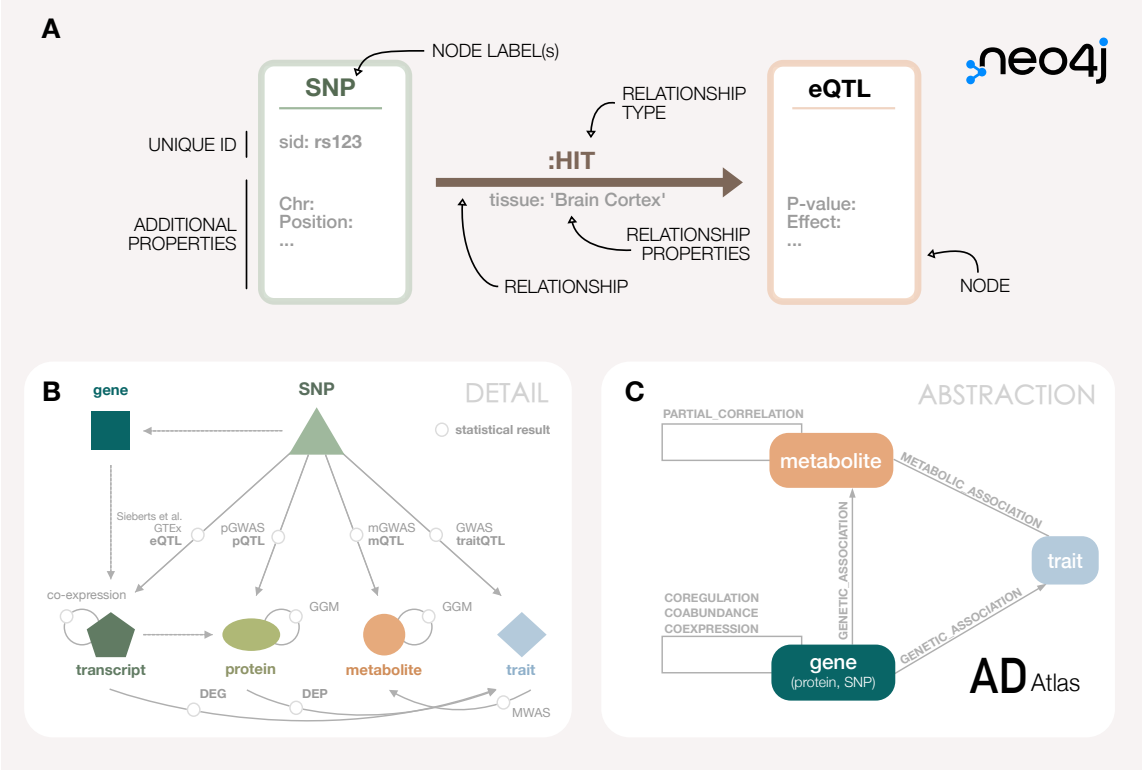


Figure 4.2: Data model underlying the AD Atlas. (**A**) Schematic representation of the Neo4j property graph model that is used to store data. (**B**) The collection of omics data, i.e., information on measured entities (nodes) and statistical or knowledge-based relationships between them (edges), are stored in a comprehensive graph structure. The data model depicted is simplified for clarity. (**C**) To enable efficient data access and reduce the complexity of the resulting networks, an abstracted data view is constructed (Figure 4.3) by projecting SNPs, transcripts, and proteins onto genes and consolidating metabolites across analytical platforms (Figure 4.4). Multiple edges connecting the same entities are summarized, retaining as much detailed information as possible in edge annotations.

Phase 1: Data collection and integration

Known biological relationships between omics layers were taken from public databases. This includes gene-transcript-protein mappings from Ensembl [332] and mappings between SNPs and genes from SNiPA [148]. SNPs, genes, transcripts, and proteins were stored in the Neo4j database as individual nodes, and relationships were added as edges between them. Furthermore, metabolites (and their corresponding meta-information, i.e., biochemical name, pathway annotations, and additional identifiers) measured across different metabolomics platforms were also collected and stored as nodes in the database. Large-scale quantitative data from population-based studies were then used to establish data-driven relationships within (e.g., tissue-specific gene co-expression) and across omics (e.g., eQTLs or mQTLs) layers. To identify entities within this network that are relevant to AD, we used large-scale association data for AD yielded in case-control and AD biomarker GWAS, MWAS, data on differentially expressed genes and differentially abundant proteins, and brain region-specific gene co-expression and protein co-abundance. Summary statistics

of these analyses were either downloaded from publicly available data repositories (listed in Section 4.1.3), supplementary data, or calculated in additional analysis steps (described in Section 4.2.3). A summary of data types and their respective sources is listed in Table 4.3.

Data format and integration. All data were transformed into a standard data format, where one (or more) comma-separated values (CSV) files were constructed for each study, containing one relationship per row (e.g., gene 1, gene 2, p-value, etc.) and an additional CSV file containing information on the data source (e.g., publication, author, year, etc.). Data integration across studies/sources is performed by overlaying the inferred knowledge-based and data-driven relationships. Data-driven associations between two entities are modeled through an intermediate node containing the respective summary statistic (seen in Figure 4.2 B). Additionally, every statistical results node is connected to a node of type *source*, enabling a highly granular view of the data and preserving information, for example, on the cohort, publication, and sample type.

ID harmonization. To facilitate integration through overlap, each biological entity, i.e., gene, transcript, protein, SNP and metabolite, was mapped to a unique identifier, either using the mapping provided by the data source or through manual curation. The unique identifier for SNPs is defined as their rsID [347], while genes, transcripts, and proteins are identified by their respective Ensembl identifiers, and measured metabolites are identified by their platform-specific ID. Additional identifiers, including biochemical metabolite names, gene symbols, and UniProt [348] identifiers, have also been annotated but are not required to be unique. AD-specific (endo-)phenotypes from different studies (GWAS,

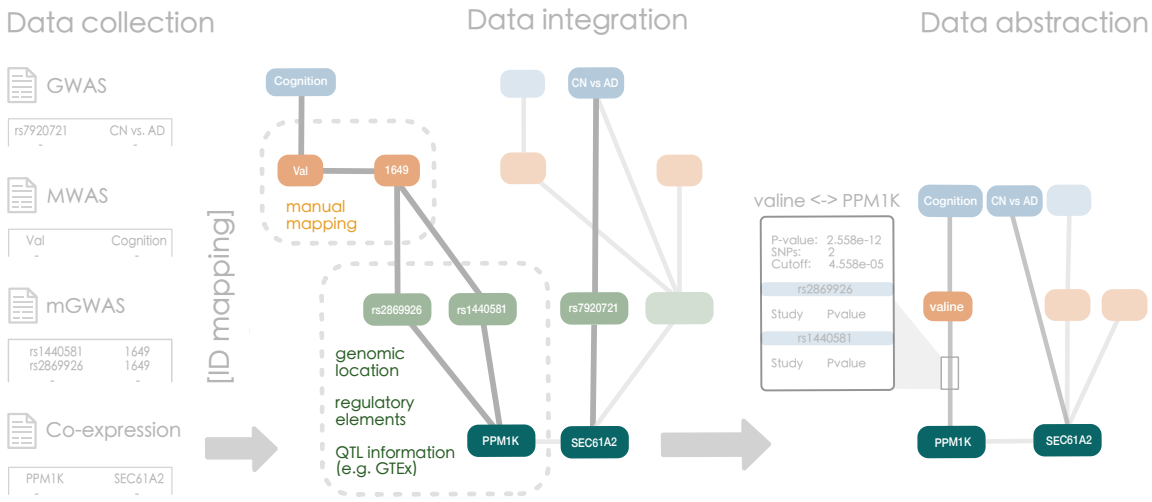


Figure 4.3: Overview of the integration pipeline and data abstraction. First, statistical results are collected from different studies and analyses ('Data collection') and subsequently stored in comprehensive network structures that consist of different node types (e.g., trait, gene, SNP, metabolite), which are connected by relationships inferred from the collected datasets ('Data integration'). Lastly, the complex data model is simplified and abstracted by manual mapping of entities across platforms and QTL-based integration to establish direct links between omics layers ('Data abstraction').

MWAS) were harmonized through manual curation and are listed in Supplementary Table 8.2. Further information regarding each phenotype is given in Supplementary Table 8.3.

Phase 2: Summary and abstraction

The resulting heterogeneous omics network is highly complex. To make data requests more efficient and allow a more intuitive (visual) interpretation of the results, the data model was projected into a simpler network view which is accessible via the AD Atlas user interface (www.adatlas.org). Figure 4.3 schematically shows an example of the abstraction pipeline described further in the following.

Projection onto genes. SNPs, transcripts and encoded proteins are projected onto genes using information from Ensembl [332], GTEx [143], and SNiPA [148]. Proteins are mapped to genes via transcripts using the mapping provided by Ensembl. For projecting SNPs to genes, we use (i) the genomic location of the SNPs: if a variant is either directly located within the gene body or 2.5kb up- or downstream, it is assigned to the respective gene(s); if a variant is located in gene-associated regulatory elements (promoters or enhancers from ENCODE [335] and FANTOM5 [336]), it is assigned to the respectively associated genes; and (ii) association data: a SNP is further assigned to a gene for which significant eQTL and pQTL associations exist. Consequently, one-to-many SNP-to-gene mappings are possible. Therefore, GWAS signals are not limited to a single gene for each

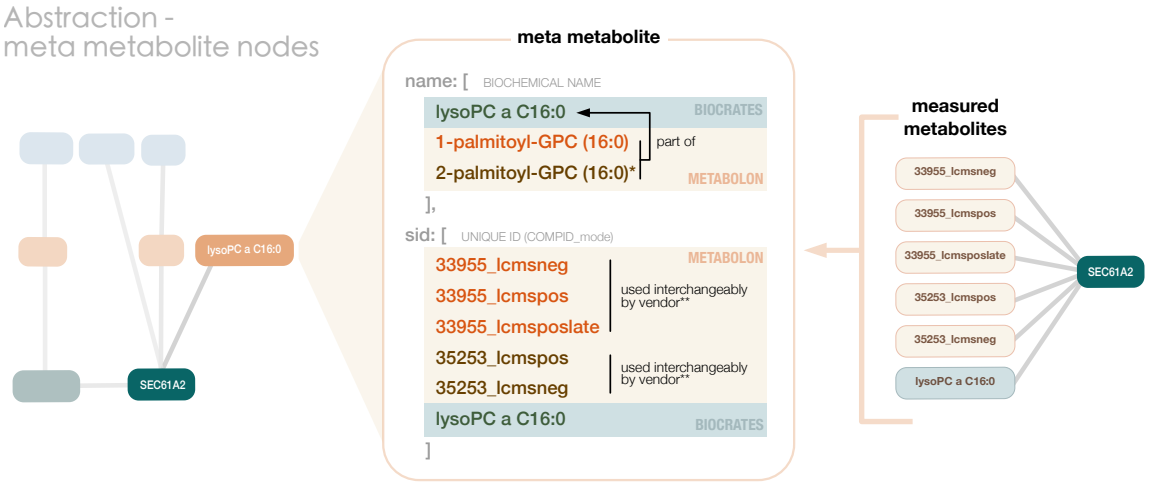


Figure 4.4: Detailed example of a meta metabolite. The meta metabolite lysoPC a C16:0 is a consolidation of six different metabolite measurements (three unique measurements, up to three mass spectrometry (MS) modes) that have been measured on two different metabolomics platforms (Metabolon and Biocrates). As these two vendors provide different levels of resolution, we map higher resolved metabolites to the lowest given level through manual cross-platform mapping. For example, 1-palmitoyl-GPC (16:0) and 2-palmitoyl-GPC (16:0)* are both constituents of lysoPC a C16:0, which is the sum of the two. Metabolites measured by Metabolon with the same compound ID but differing measurement modes are also summarized, as the vendor measures some compounds across several MS modes and reports the measurement from the platform/mode showing the best performance for the respective compound (determined on a per-study basis; thus assumed as being interchangeable**).

locus but mirror shared genetic and epigenetic regulation to provide a comprehensive representation of each locus. The information used for this projection is partly extracted from SNiPA [148] and partly comes from the GTEx [143], or Sieberts et al. [327] studies. Additionally, genes displayed in the current version of the AD Atlas have been limited to protein-coding genes. The resulting meta-gene nodes contain all relationships from the individual omics layers (e.g., from transcriptomics and proteomics analyses).

Metabolite consolidation. Metabolites that were measured on more than one platform were consolidated using manual mappings between platform-specific identifiers and information on identified unknown metabolites to construct so-called 'meta metabolite' nodes (statistics given in Supplementary Tables 8.4 and 8.5). Furthermore, metabolites measured by the Metabolon platform, which were annotated with the same biochemical name and chemical ID (but differing compound ID), were also merged. A detailed example of metabolite consolidation is given in Figure 4.4.

Edge summary and significance thresholds. Statistical associations between biological entities (edges) were summarized while retaining all information on the underlying SNPs and proteins. This information is stored in comprehensive edge labels. Edges are further pre-filtered using study-specific, gene-wise, or genome-wide significance thresholds. An in-depth discussion regarding this decision is presented in Chapter 6. Genetic associations with AD traits (trait QTLs) and metabolites (mQTLs) can be filtered using either a genome-wide ($P \leq 5 \times 10^{-8}$) or a Bonferroni-like gene-wise cutoff. The gene-wise cutoff for a $gene_A$ is defined by:

$$\text{p-value} \leq \frac{0.05}{\#\text{SNPs}_{gene_A}} \tag{4.1}$$

where $\#\text{SNPs}_{gene_A}$ is the number of SNPs that have been annotated to $gene_A$ through SNiPA [148]. A comprehensive list of study-specific significance thresholds is given in Table 4.3.

4.4 Network-based multi-omics analyses

The following provides an overview of the downstream analyses conducted to evaluate and further explore the comprehensive collection of multi-omics associations (Figure 4.5). They can broadly be categorized into global analysis, i.e., methods that take the whole network (structure) into account (Section 4.4.1) and local analysis, i.e., hypothesis-guided generation and exploration of context-specific subnetworks (Sections 4.4.2 and 4.4.3). All analyses were performed using the abstracted data model, as seen in Figure 4.2 C.

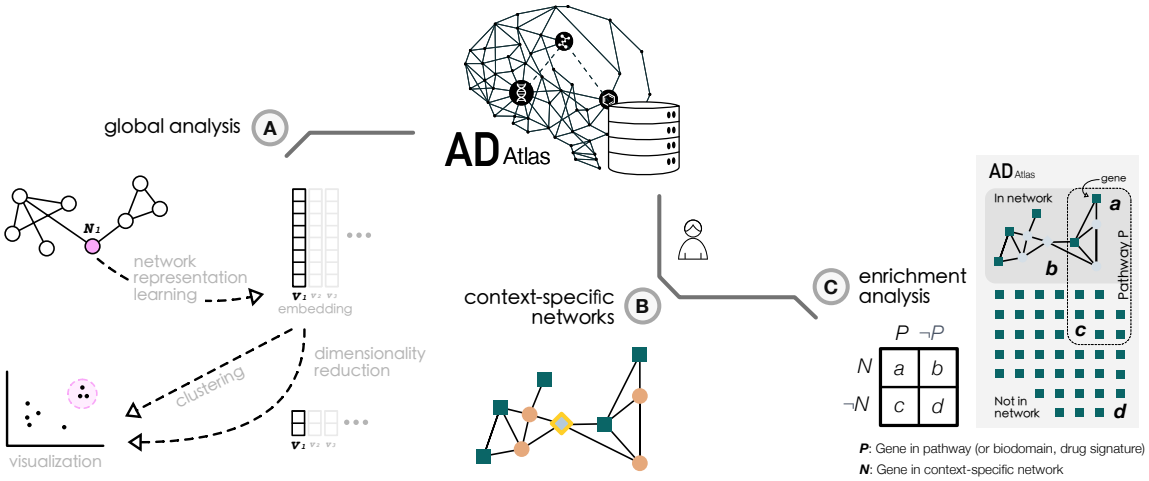


Figure 4.5: Network-based multi-omics analyses using the AD Atlas. The comprehensive collection of correlational data can either be analyzed globally (A) or locally (B). (**A**) The heterogeneous network is transformed into a simple graph (i.e., undirected, and without self-loops or multiple edges) and embedded into a lower-dimensional, latent-space representation using network representation learning algorithms (DeepWalk [349, 350]). This vector representation of the network can then be clustered using, for example, hierarchical clustering, and visualized in 2D by further dimensionality reduction (Uniform Manifold Approximation and Projection (UMAP) [351]). (**B**) Annotation of (a set of) biological entities (metabolites, genes) or AD-related traits with significantly associated entities provides a multi-omics view. To make these multi-omics networks context-specific, i.e., tailored to a specific research question, the results can be restricted to a specific brain region or sample type and expanded to the functional neighborhood through gene co-expression, protein co-abundance, and metabolite partial correlation networks. (**C**) Once built, networks can be subjected to enrichment analysis, for example, to functionally characterize the genes or metabolites present in the network.

4.4.1 Network representation learning and visualization of global network structure

Network representation learning on the global AD Atlas network (Figure 4.5 A) was performed using a slightly modified version of the node embedding framework implemented in the *GeneWalk* Python package [350]. Details of this approach can be found elsewhere [349, 350]. Briefly, vector representations of nodes are learned by sampling unbiased random walks over the network and using neighboring node pairs from these walk sequences as input-output training sets for a fully connected neural network. The package also enables the generation and embedding of randomized networks that retain the same degree distribution which was used to compare structures seen in the AD Atlas (Supplementary Figure 8.3 A). A simplified graph was abstracted from the AD Atlas, i.e. an undirected graph that includes no self-loops and multiple edges between two nodes are only represented by one edge. Furthermore, no distinction was made between the different node and edge types as this information cannot be taken into account by the algorithm. As we were interested to see if the global structure of the network provides relevant biological information, we excluded AD phenotype nodes from the network. A brain-specific network was constructed by restricting the sample type of all edges to 'brain' (protein

co-abundance, metabolite partial correlation, metabolite-trait associations, and mQTL data). A genome-wide significance filter ($P \leq 5 \times 10^{-8}$) was applied to genetic associations. As the SNP-to-gene annotation underlying QTL-based edges incorporates eQTL information, we did not include co-regulation edges in the network to avoid overweighting this data modality (Supplementary Figure 8.3 C). The node degree distribution for this brain-specific network can be seen in Figure 8.2. GeneWalk was run setting the window size (context surrounding each node that is regarded for embedding) to 2 and the dimension of the vector representations (embeddings) to 89 (–dim_rep=89). The appropriate embedding dimension was determined by using the approach described in [352]. As this method uses the embedding algorithm Node2Vec [353] we set the parameters p and q both to 1 and the *window_size* to 2 to make the embedding algorithm similar to the embedding algorithm used by GeneWalk (DeepWalk [349]). For visualization, the resulting node vectors were projected to two dimensions using UMAP [351] as implemented in the Python library *umap* (settings: n_neighbors=15, min_dist=0.1, metric: 'cosine'). Hierarchical clustering and visualization as a dendrogram were performed using the R packages *stats* (v4.2.0) and *dendextend* (v1.16.0). The 89-dimensional node vectors were used as input and the method 'ward.D2' with Euclidean distance was used for clustering.

4.4.2 Generation of context-specific networks

A context-specific molecular network (Figure 4.5 B), in the context of this work, refers to a network that has been built surrounding a (set of) gene(s), metabolite(s), or AD-related phenotype(s) of interest through the addition of pairwise association data. The AD Atlas enables the dynamic annotation of a collection of input entities in the context of the research question at hand, which includes additional filtering options and possibilities to expand the initial input set. Settings that were used in the scope of this work and that have been implemented through the AD Atlas user interface are explained in more detail in the following.

Molecular subnetwork generation. Networks can be built surrounding one or more input entities. The AD Atlas provides three distinct entry points:

(i) Trait- or meta-trait-centric subnetworks
The user specifies a trait or collection of traits as input. Genes and metabolites directly associated with the traits of interest are added using data from genome- and metabolome-wide association studies. Next, these entities are annotated with associated genes and metabolites (through mQTL data) and finally, all associated traits are added for this final set of genes and metabolites, again using data from genome- and metabolome-wide association studies. All relationships between these entities that meet the user-specified criteria, including intra-omics links, are included

in the returned network. Here, meta-traits are pre-defined collections of traits loosely following the A-T-(N)-(C) schema [251] (see Supplementary Table 8.2).

(ii) Gene-centric subnetworks
The user provides a gene or a set of genes as input (provided as gene symbol or Ensembl gene ID). Traits and metabolites that are directly associated with the genes of interest are added using data from GWAS with AD-related traits and GWAS with metabolic traits. Lastly, all relationships between these entities that meet the user-specified criteria are added to the resulting network.

(iii) Metabolite- or Pathway-centric subnetworks
The user provides a metabolite or a set of metabolites as input (provided as a biochemical name). Traits and genes that are associated with the metabolite of interest are added using data from MWAS with AD-related traits and mQTL data generated in GWAS with metabolic traits. Lastly, all relationships between these entities that meet the user-specified criteria are added to the resulting network. Here, pathways are predefined sets of metabolites, currently defined by pathway annotations (super-/sub-pathways and metabolite classification) as provided by the metabolomics lab or platform vendor that generated the data.

Network expansion. Prior to building the network, the initial set of input genes or metabolites can be expanded to include the 1-step or 2-step functional neighbors. For gene-centric networks (input: gene(s)) these neighbors can be defined via gene co-expression, protein co-abundance, and/or eQTL co-regulation networks. For metabolite-centric networks (input: metabolite(s)), expansion is performed using partial correlation networks (GGMs). Context filtering is applied for this step if the user has specified a sample type, tissue, or brain region. Once the selection of input nodes has been expanded, they are annotated with associated entities as described above.

Significance threshold. Inferred links between and within omics layers are filtered by applying basic significance thresholds. These are predominantly study-specific and can be found in Table 4.3. For genetic associations, that is links between genes and metabolites (inferred from mGWAS), and genes and traits (inferred from GWAS) users can select either a gene-wise (Equation 4.1) or genome-wide ($P \leq 5 \times 10^{-8}$) threshold. This gene-wise significant threshold is less stringent than the genome-wide cutoff and will yield more results.

Sample type filtering. Edges can be filtered by sample type, tissue, or brain region. Currently, links established through tissue-specific eQTL analysis (COREGULATION) taken from GTEx [143] and large-scale analysis of cortical cis-eQTLs by Sieberts et al. [327] can be filtered by 49 tissues, of which 13 are brain-specific. Furthermore, gene

co-expression networks (COEXPRESSION) were constructed for seven different brain regions by which these links can be filtered. GENETIC_ASSOCIATION (gene-metabolite), METABOLIC_ASSOCIATION, COABUNDACE, and PARTIAL_CORRELATION edges can be filtered by sample type, which currently includes brain, plasma, serum, and urine. If nothing is specified, this setting defaults to no filtering. The filtering, when specified, is also applied to the network expansion step.

Visual overlay of differential expression data. To assess the extent and direction of dysregulation in AD, data from differential analysis in AD cohorts can be visually projected onto the built multi-omics networks. This includes information about genes that are differentially expressed in the brain [328], called differentially expressed genes (DEGs), and proteins that are differentially abundant in the brain [329, 330], called differentially abundant proteins (DEPs). Up-regulation, down-regulation, or non-regulation at the transcriptional level is indicated by coloring the gene nodes in the network in red, blue, and light green, respectively. Differential protein abundance is indicated by coloring the edge of the gene nodes, using the same color coding as for genes.

4.4.3 Functional and statistical network assessment

Multi-omics context networks can be subjected to various downstream analyses. In this work, over-representation analysis (ORA) using gene annotations and drug signatures was performed to better understand the biological context of selected context networks and their ability to identify candidate drugs for repositioning. In addition, empirical p-values were calculated for these examples to assess the statistical significance of these networks in terms of network statistics, such as the extent of disease association and their overall background distribution. Further details are provided below.

Functional enrichment analysis

Enrichment analysis, also referred to as ORA, is used to identify annotation terms (e.g., biological pathways, metabolite classes, or disease gene sets) that are more frequently assigned to entities in a list of interest than expected by chance. Here, these entities of interest consist of the genes or metabolites contained in a context-specific network. Over-representation of a specific annotation term is tested by constructing a contingency table (schematically depicted in Figure 4.5 C) and conducting a one-sided Fisher's exact test [110] with subsequent correction for multiple testing. All annotated entities present in the AD Atlas are considered as background, where possible.

To perform enrichment analysis of the generated multi-omics context networks, the AD Atlas provides a variety of different tools:

- **Gene set enrichment** of the generated subnetworks can be performed using the R package *enrichR* (v3.0) [354], which enables the analysis of a variety of different gene sets, including drug perturbation signatures and biochemical pathways, via the Enrichr webservice (https://maayanlab.cloud/Enrichr/). The Benjamini-Hochberg method is used to account for multiple testing and the corrected p-values are reported as q-value. Enrichr does not allow the definition of a custom background list. EnrichR gene set libraries used throughout this work are the following: *Drug_Perturbations_from_GEO_2014* (n=701 gene lists) - where each Food and Drug Administration (FDA) approved drug is a list consisting of the associated differentially expressed genes as identified by Gene Expression Omnibus (GEO) [355] experiments; *Drug_Perturbations_from_GEO_down/up* (n=906) – manually extracted drug signatures created from RNA-Seq data in GEO as part of a crowdfunding initiative; *Reactome_2016* (n=1530) – pathway gene sets taken from the Reactome pathway knowledge base [356] and *WikiPathway_2021_Human* (n=622) – pathway gene sets taken from the biological pathway database WikiPathways [357]. Additionally, enrichment can be performed using the R package *gprofiler2* (v0.2.0) [358], which provides an interface to the gene list functional profiling toolset, g:Profiler (https://biit.cs.ut.ee/gprofiler/gost). The correction method applied for multiple testing can be chosen by the user and the background set is defined as all protein-coding genes included in the AD Atlas.

- **GO term enrichment** is provided via the R package *topGO* (v2.38.1) [359]. The Gene Ontology (GO) is a structured representation of gene function classification, separated into three sub-ontologies; biological process, molecular function, and cellular component. As a background list, all protein-coding genes included in the AD Atlas are used.

- **Metabolic pathway enrichment** analysis for metabolites included in the constructed molecular subnetworks is performed using platform-specific annotations of metabolites into classes and super-/sub-pathways. Contingency tables are calculated and then tested using Fisher's exact test (function *fisher.test* with alternative = 'greater' from the R package stats). Multiple testing correction is performed using *p.adjust* (method = 'fdr'). As background, we use all metabolites for which pathway information is available.

- **Alzheimer's disease biological domain enrichment** analysis is performed using manually curated AD-relevant gene function annotations. To date, this ongoing curation effort has categorized 54% of all genes by summarizing over 8,200 GO terms into 16 AD-related biological domains: Synaptic Function, Immune Response, Endolysosomal Trafficking, Structural Stabilization, RNA Spliceosome, Myelination, Vascular Function, Mitochondria and Metabolism, Autophagy, Apoptosis, Epigenetics, Oxidative Stress, Lipid Metabolism, APP Processing, Tau homeostasis, Proteostasis.

Further information on the approach and the complete mapping of genes and GO terms to biological domains is available at https://www.synapse.org/#!Synapse:syn26529354. The ORA is performed analogously to the above-described metabolic pathway enrichment analysis. As background, we use all genes that have been annotated with biological domains.

- **Enrichment of differentially expressed genes/proteins** is performed analogously to the above-described metabolic pathway and biodomain enrichment analysis. Here, up-regulated, down-regulated and unchanged genes/proteins are treated as individual gene annotation sets. As background, we use all genes/proteins that have been tested for differential expression.

Statistical significance of network properties

To compare the network statistics of a hypothesis-guided network to the AD Atlas background distribution, we estimated empirical p-values by generating further networks using the same input parameters (e.g., tissue filtering, significance threshold) but with random entities as starting points. The resulting p-values give an idea of how likely it is that the observed network statistics (e.g., a high number of up-regulated genes) may have arisen by chance. Given the observed (network) statistic t and a set X which contains the (network) statistics from n randomly generated networks $X = \{x_1, x_2, ..., x_n\}$, the empirical p-value can be approximated by the fraction of observations that are equal to or more extreme than t which is given by

$$P_{emp} \approx F_n(t) = \frac{\sum_{i=1}^{n} \mathbb{1}(x_i \geq t)}{n} \tag{4.2}$$

where $\mathbb{1}(\cdot)$ is the indicator function. Consequently, the smallest possible p-value achievable before reaching zero, i.e., the observed statistic is larger than any statistic from a randomly generated network, is determined by the number of generated networks [360] as $\frac{1}{n}$. This should be taken into account when interpreting the approximated empirical p-values.

Empirical p-values were approximated using the above-described Equation 4.2 for the following network properties: network size – the number of nodes (genes, metabolites, and traits in the network); the proportion of entities (genes or metabolites) significantly associated with AD-related traits (number of trait-associated entities in network / total number of entities in the network); proportion of genes significantly associated with AD-related traits (number of trait associated genes in network / total number of genes in network); proportion of metabolites significantly associated with AD-related traits – analogous to the previous; proportion of DEGs (DEGs in network / all entities tested for differential expression in network); proportion of DEPs – analogous to DEGs. The empirical p-values were estimated from the background distribution of the proposed statistics.

For each showcase, we generated 1000 random networks using the same input parameters but replacing the hypothesis-guided input genes/metabolites with randomly selected genes/metabolites. For the trait-centric network, the genes and metabolites associated with the input trait were replaced by random entities of the respective type and then subjected to network generation by annotation of direct associations (GENETIC_ASSOCIATION and METABOLIC_ASSOCIATION edges).

4.5 Technical framework of the AD Atlas

The AD Atlas resource backend was built using the statistical computing language R and the graph database management system Neo4j (community v4.4.11). After pre-processing and adding additional metadata (e.g., information on the source and sample type), data files are deposited in specific folder structures. Loader R scripts are used to fill the

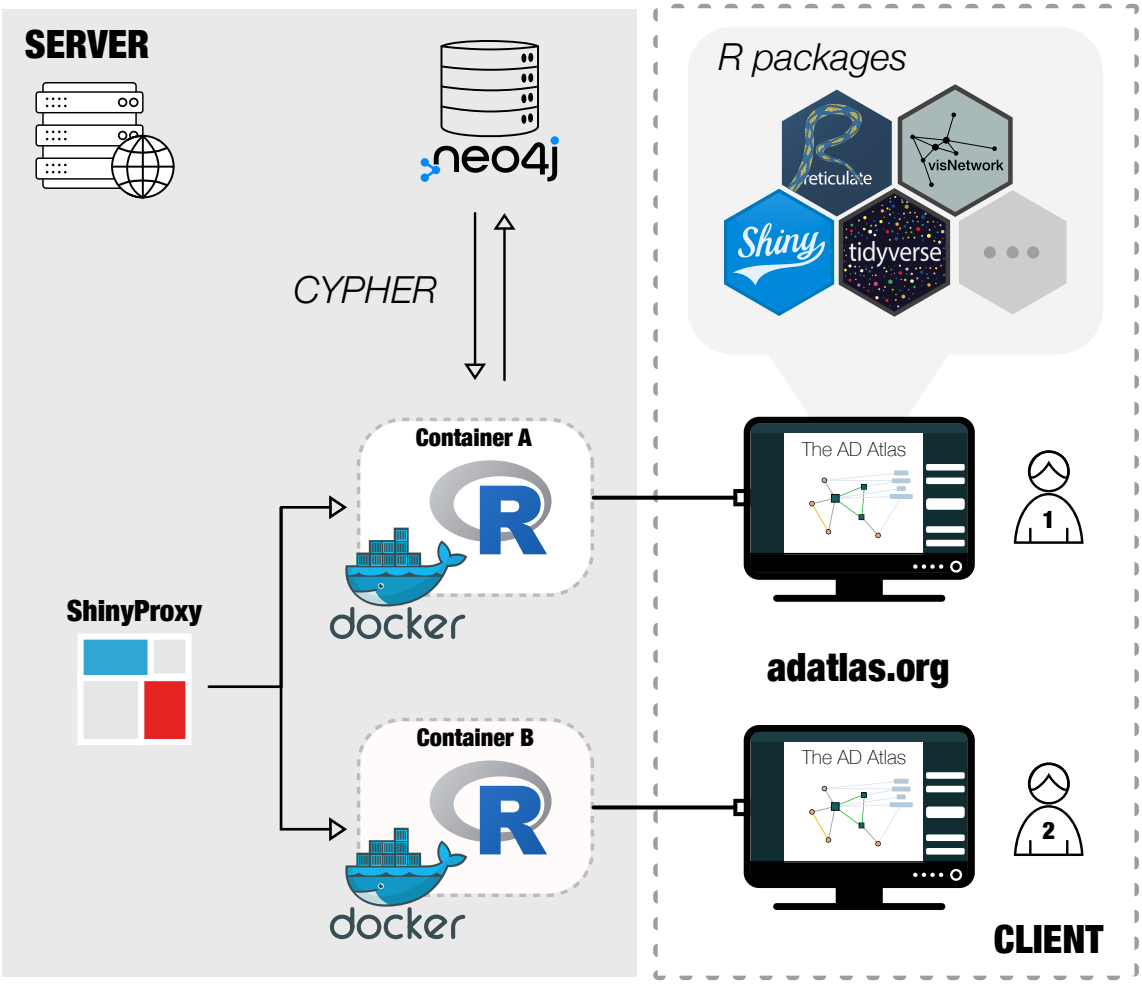


Figure 4.6: IT-architecture of the AD Atlas. The AD Atlas frontend was implemented as an R ShinyApp that communicates with the Neo4j backend of the AD Atlas via cypher requests through the Neo4j python driver *neo4j*. The app is deployed using ShinyProxy which provides an isolated app environment for each user session.

database (DB), which natively stores data in accessible network structures. Biological entities and their respective annotations are represented as nodes and the inferred significant relationships between them as edges. Import requests are sent to the DB using the graph query language Cypher. Communication between Neo4j and R is established through the official Neo4j python driver *neo4j* (https://github.com/neo4j/neo4j-python-driver) and the R-to-python interface package *reticulate* (v1.17). Interactive networks are visualized using *visNetwork* (v2.0.9). The AD Atlas can be rebuilt from scratch. This ensures full reproducibility and consistency throughout time.

To enable access to this multi-omics resource and provide additional analysis tools, we implemented a network- and web-based user interface using an R ShinyApp. Due to R's single-threaded nature, the AD Atlas frontend is deployed using the open-source solution ShinyProxy (v2.3.1_amd64.deb). Here, isolated app environments are provided for each session (user) using Java and Docker containers. A general overview of the IT architecture of the AD Atlas can be seen in Figure 4.6. Context-specific subnetworks are constructed by translating the user-specified search parameters into a cypher query that is subsequently sent to the Neo4j graph database to retrieve relevant information. The resulting data matrices, one with entity (node) information and one containing information on the relationships (edges), are processed using R and provided to the AD Atlas frontend. A detailed example query is shown in Supplementary Figure 8.1. All tools used are publicly available and free for academic use. In particular, we used the following:

Data storage

- Neo4j: Version: v4.4.11 (community). https://neo4j.com
- Docker: Version 1.13.1, Build cccb291/1.13.1. https://www.docker.com
- ShinyProxy: Version 2.3.1_amd64.deb. https://www.shinyproxy.io/

User interface - R packages

- shiny: Version: 1.6.0. Web application framework for R.
- shinyjs: Version: 2.0.0. Improves the user experience of your shiny apps.
- scroller: Version: 0.1.1. Scroll to any element in Shiny.
- shinyEffects: Version: 0.2.0. Customize your web apps with fancy effects.
- shinybusy: Version: 0.3.1. Busy indicator for 'Shiny' applications.
- shinydashboard: Version: 0.7.1. Create dashboards with 'Shiny'.
- shinydashboardPlus: Version: 0.7.0. Add more 'AdminLTE2' components to 'shinydashboard'.
- shinyWidgets: Version: 0.6.0. Custom inputs widgets for shiny.
- shinycustomloader: Version: 0.9.0. Custom loader for shiny outputs.
- particlesjs: Version: 0.1.1 Beautiful background for 'Shiny' Applications and 'Rmarkdown' Documents

Data visualization - R packages After processing the data frames returned by the user-defined cypher query, the igraph function *degree* is used to extract the node degrees (used for scaling of node size) and the functions *layout_with_drl*, *layout_with_graphopt* from the igraph package or the qgraph package function *qgraph.layout.fruchtermanreingold* is used to precalculate the network layout. Depending on user input, this is either done for the whole network (Include traits in layout=yes) or excluding trait nodes and arranging these separately, to the right of the layout (Include traits in layout=no). Finally, the visualization of interactive networks is performed using the R package *VisNetwork*. Network representation as hive plots is rendered with the R package *HiveR*, 3D networks are rendered with *threejs*, and interactive bar- and donut-plots of the network statistics are realized using *plotly*. The package *DT* is used to create HTML table widgets that allow users to sort, search and download tables.

- VisNetwork: Version: 2.0.9. Network visualization using 'vis.js' library.
- qgraph: Version 1.6.5. Graph plotting methods and psychometric data visualization.
- igraph: Version 1.2.4.1. Network analysis and visualization.
- threejs: Version 0.3.3. Interactive 3D scatter plots, networks, and globes.
- plotly: Version 4.9.4.1. Creative interactive web graphics via 'plotly.js'.
- DT: Version: 0.16. A wrapper of the JavaScript library 'DataTables'.
- HiveR: Version: 0.3.63. 2D and 3D Hive plots for R.

Enrichment analysis – R packages

- topGO: Version: 2.38.1. Enrichment analysis for Gene Ontology.
- enrichR: Version: 3.0. Provides an R interface to 'Enrichr'.
- gprofiler2: Version: 0.2.1 Provides an R interface to 'g:Profiler' toolset.

An online molecular atlas of Alzheimer's disease 5

Results presented in this Chapter have been used to prepare the following two first-author manuscripts (*peer-reviewed journal submission in preparation*):

1. M. A. Wörheide, J. Krumsiek, S. Nataf, K. Nho, A. K. Greenwood, J. C. Wiley, T. Wu, K. Huynh, P. Weinisch, W. Römisch-Margl, N. Lehner, The AMP-AD Consortium, The Alzheimer's Disease Neuroimaging Initiative, The Alzheimer's Disease Metabolomics Consortium, J. Baumbach, P. J. Meikle, A. J. Saykin, P. Murali Doraiswamy, C. van Duijn, K. Suhre, R. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"An online molecular atlas of Alzheimer's disease"**, *submission in preparation.* Preprint available on medRxiv, doi: 10.1101/2021.09.14.21263565.
2. M. A. Wörheide, J. Krumsiek, S. Nataf, K. Nho, A. K. Greenwood, J. C. Wiley, T. Wu, K. Huynh, P. Weinisch, W. Römisch-Margl, N. Lehner, The AMP-AD Consortium, The Alzheimer's Disease Neuroimaging Initiative, The Alzheimer's Disease Metabolomics Consortium, J. Baumbach, P. J. Meikle, A. J. Saykin, P. Murali Doraiswamy, C. van Duijn, K. Suhre, R. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"Utilizing multi-omics context networks to explore molecular hypotheses in Alzheimer's disease"**, *submission in preparation.* Preprint available on medRxiv, doi: 10.1101/2021.09.14.21263565.

Contributions of the lead author for both manuscripts were as follows:

- implementation of the project
- data analysis and resource development (backend and frontend)
- drafting the manuscript
- design and implementation of all figures

Critical revision of the manuscripts and final approval was given by all.

5.1 Background

Late-onset Alzheimer's disease (AD) is a progressive neurodegenerative disorder, for which there is currently no cure or preventive therapy and only modestly effective symptomatic treatments [361]. The failures of hundreds of trials of disease-modifying therapeutics, including several phase III trials targeting amyloid-beta (Aβ), and availability of only two Food and Drug Administration (FDA)-approved anti-amyloid compounds [305, 306, 362]

highlight our incomplete understanding of both the cause of AD and the mechanisms of cognitive failure [363]. AD is a multifactorial disease with a long prodromal period as well as substantial heterogeneity in both risk profiles and clinical/pathological presentation. It is linked to all molecular layers from genetic and epigenetic variation through transcriptional changes to altered abundances of proteins and metabolites, which interact in complex networks [364]. Hence, AD is best viewed as a complex alteration in many molecular readouts, which can be seen as a shift of a multi-molecular network from a "normal" to a perturbed state.

Despite significant advances in the study of AD and related dementias, there are many challenges remaining as recently highlighted in the 2021 National Institutes of Health (NIH) AD research summit. One of the most prominent missing pieces are robust and reliable biomarkers for both diagnosis and therapeutic intervention that are embedded in the context of multi-level molecular changes observed in AD and are evaluated in an open, rigorous, and reproducible manner. NIH's Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) (https://www.nia.nih.gov/research/amp-ad; [4]) program is working towards this goal through the generation and examination of diverse data including multi-omics profiling of different modalities and across relevant tissues, where all data generated through the AMP-AD initiative is rapidly shared through the AD Knowledge Portal (https://adknowledgeportal.org; [325]).

The AD Knowledge Portal's Agora Platform (https://agora.adknowledgeportal.org/) provides interactive visualizations designed to support the evaluation of data from RNA-seq, proteomics, and metabolomics studies on the single target level. However, a user-friendly analytical tool that incorporates single biological entities into their multi-omics context has so far been missing. To this end, networks offer an intuitive framework to integrate and store densely connected biomedical data, making them an attractive data structure for multi-omics integration efforts [7]. Heterogeneous networks, which consist of multiple types of nodes (e.g. metabolites, genes, and phenotypes) and edges (e.g. partial correlation of metabolites, gene co-expression), have been particularly useful to describe the complex interplay within and between biological domains [365]. Such network-based multi-omics approaches have the potential to ultimately construct comprehensive and largely bias-free models of AD that can guide the identification and prioritization of potential therapeutic targets and drug repositioning candidates [239, 304] as well as inform novel hypotheses that can be tested in follow-up experiments.

The following Chapter presents the AD Atlas, a network-based data integration resource for investigating AD, its biomarkers, and associated (endo-)phenotypes in a multi-omics context. The AD Atlas integrates data from more than 25 studies using an extended quantitative trait loci (QTL)-based integration strategy combined with a composite network approach [10]. Based on data from knowledge bases and healthy cohorts, we constructed a generalized, disease-independent high-quality framework of intra- (e.g. gene-

Table 5.1: Data compiled in the AD Atlas (simplified data view).

		$N_{entities}$	Ref.
Nodes	AD-related **meta-trait**	10	-
	AD-related **traits**	67 43 unique	-
	Genes (protein-coding)	20,363	
	with differential gene expression data	14,731	[328]
	with differential protein abundance data	7,867	[329] [330]
	Metabolites	1,328	-
Edges	**Genetic associations with AD phenotypes** traitQTLs	12,361	[278] [338] [339] [340] [341] [275] [274] [272] [342] [273] This study*
	Genetic associations with metabolic traits mQTLs; sample types (n=3)	165,719	[29] [25] [26] [24] [337] This study*
	Metabolic associations with AD phenotypes mWAS; sample types (n=2)	1,018	[281] [280] [245] [263]
	Gene co-expression data brain regions (n=7)	232,516	[328]
	Genetic co-regulation eQTLs; tissues (n=49)	493,117	[327] [143]
	Protein co-abundance data partial correlation-based; sample types (n=2)	73,296	[146] [329] [330]
	Metabolic pathways partial correlation-based; sample types (n=2)	1,163	[168] [263] This study*

*For details on datasets generated in the scope of this work, please refer to Chapter 4 Section 4.2.3.

gene) and inter-omics (e.g. metabolite-gene) relationships. Using large-scale association data of AD – including data from AMP-AD, NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS), and other large studies and consortium efforts – this framework was then transformed into an integrated multi-omics knowledge base for markers of AD. The resulting comprehensive catalog of multi-omics relationships, stored using the graph-based database Neo4j, provides disease-relevant information on over 20,000 protein-coding genes, 8,000 proteins, and 1,000 metabolites as well as associated genetic variants. Lastly, we developed a publicly available network- and web-based user interface featuring several data analysis tools (www.adatlas.org) to enable access to these complex data independent of in-house bioinformatics capacities. The AD Atlas allows users to construct, expand and explore context- and tissue-specific molecular subnetworks surrounding either individual or multiple entities of interest. In the following, we demonstrate the utility of this resource

to generate disease-relevant insights using a variety of showcases ranging from hypothesis-driven to hypothesis-generating analyses. Important terms and concepts used throughout this Chapter are defined in Box 5.1.

Box 5.1: Glossary of important terms and concepts.

Step-wise multi-omics integration – Pairwise association results are overlaid and integrated to interconnect entities within and between omics layers, enabling integration across different cohorts and studies.

AD-related trait – traits that are tested for associations in genome-wide association studies (GWAS) or metabolome-wide association studies (MWAS), including cognitive measures as well as cerebrospinal fluid (CSF) and imaging biomarkers of Alzheimer's disease (AD). The term is used interchangeably with AD (endo-)phenotypes throughout this work.

Metabolic trait – measured concentrations of a metabolite which are tested for association with AD in MWAS.

SNP-to-gene mapping - For the assignment of single nucleotide polymorphisms (SNPs) to genes, we use the genomic location of a SNP (either directly within the gene body of a gene or within 2.5 kb up- or downstream), gene-associated regulatory elements (promoters or enhancers from ENCODE [335] and FANTOM5 [336]), and expression quantitative trait loci (eQTL) and protein quantitative trait loci (pQTL).

Context-specific molecular subnetwork – Multi-omics network built from pairwise association data surrounding entities of interest; associations (edges) are optionally filtered for tissue or sample type.

Annotation – Addition of statistical results (nodes and edges) to an entity or set of entities to build a context-specific molecular network. For example, annotating metabolites through metabolite quantitative trait loci (mQTL) associations with metabolic genes.

Significance threshold – Cutoff for statistical significance. Associations between two entities that are below this threshold are depicted as edges in the network.

Edge filtering – To create context- and tissue-specific networks, users can filter associations (edges) by tissue or sample type. This is currently possible for co-expression, co-regulation, genetic associations, and metabolic associations.

Edge types – categorize statistically significant associations between two entities, which are represented as an edge in the AD Atlas resource. Further details are given in Box 5.2.

5.2 Overview of the AD Atlas

The AD Atlas is a comprehensive, network-based catalog of results from large omics studies that is accessible via an interactive, network-based user interface (www.adatlas.org). It was built by inferring relationships between biological entities (e.g. metabolites, SNPs, genes) from large-scale studies, resulting in a highly complex collection of data stored in network format (Figure 4.2). To increase data accessibility and downstream interpretability, the data were summarized and extracted into a simplified data view (Figures 4.2 and 4.3). A detailed description of the integration and summarization pipeline is provided in Materials and methods (Section 4.3.2). This summarized network representation is accessible through the user interface and consists of three node types; (i) metabolites,

mapped across available metabolomics platforms where possible, (ii) genes, including information on associated transcripts, SNPs and proteins, and (iii) traits, which describe AD (endo-)phenotypes and biomarkers. Various different relationship types interconnect these entities (Figure 4.2 C and Supplementary Table 8.1). Integrating over 25 different studies and analyses, the AD Atlas includes information on 20,363 protein-coding genes, 8,396 proteins, 1,328 metabolites, and 43 unique AD-related traits. Traits include CSF and imaging biomarkers, partially with different covariate settings (adjustment for apolipoprotein E (*APOE*) genotype), stratifications or stagings for neuropathologies, amounting to 67 traits in total. Full details on the reported AD-related traits are provided in Supplementary Table (Supplementary Table 8.3). Biological entities and traits are linked by over 1.5 million relationships, representing statistical associations inferred from large-scale quantitative data from population-based cohorts and AD case-control studies. A more detailed summary of the data compiled in the AD Atlas can be seen in Table 5.1 and Box 5.2 provides further details on different edge types. Finally, to enable access to the AD Atlas we have implemented a network- and web-based user interface (Figure 5.1) which allows users to dynamically generate, explore and analyze, context-specific molecular subnetworks surrounding entities of interest i.e. genes, metabolites or AD-related (endo-)phenotypes (workflow seen in Figure 5.2).

To assess how well such an association network captures biological information with respect to both plausibility, i.e. does the multi-omics network capture known (AD) biology, and utility, i.e. can the AD Atlas provide additional, potentially AD-relevant relationships, we examine the content of the AD Atlas in the context of the available scientific literature and established knowledge bases. To this end, we utilize prior knowledge using two distinct approaches: First, we evaluate the global network structure to see if the AD Atlas as a whole captures biologically meaningful information using deep-learning-based dimensionality reduction techniques and projecting the resulting representation of the AD Atlas onto AD-related functional domains. Second, we focus on hypothesis-driven examples that explore context-specific molecular subnetworks surrounding specific genes or pathways with established links to AD pathogenesis and examine their relevance based on current AD literature. The following will shortly introduce both of these approaches.

5.2.1 Functional assessment and visualization of global network structure

Global analysis of the multi-omics network was performed by reducing the complexity of the network to enable holistic analysis and visualization in an interpretable manner. To achieve this, we applied a node embedding approach [350] in which nodes in the network are embedded in a lower-dimensional vector space using random walks while retaining structural information of the network [366]. Details of this method are provided in Chapter 4 Section 4.4.1. In the resulting embedding space, similar/proximal nodes in the original

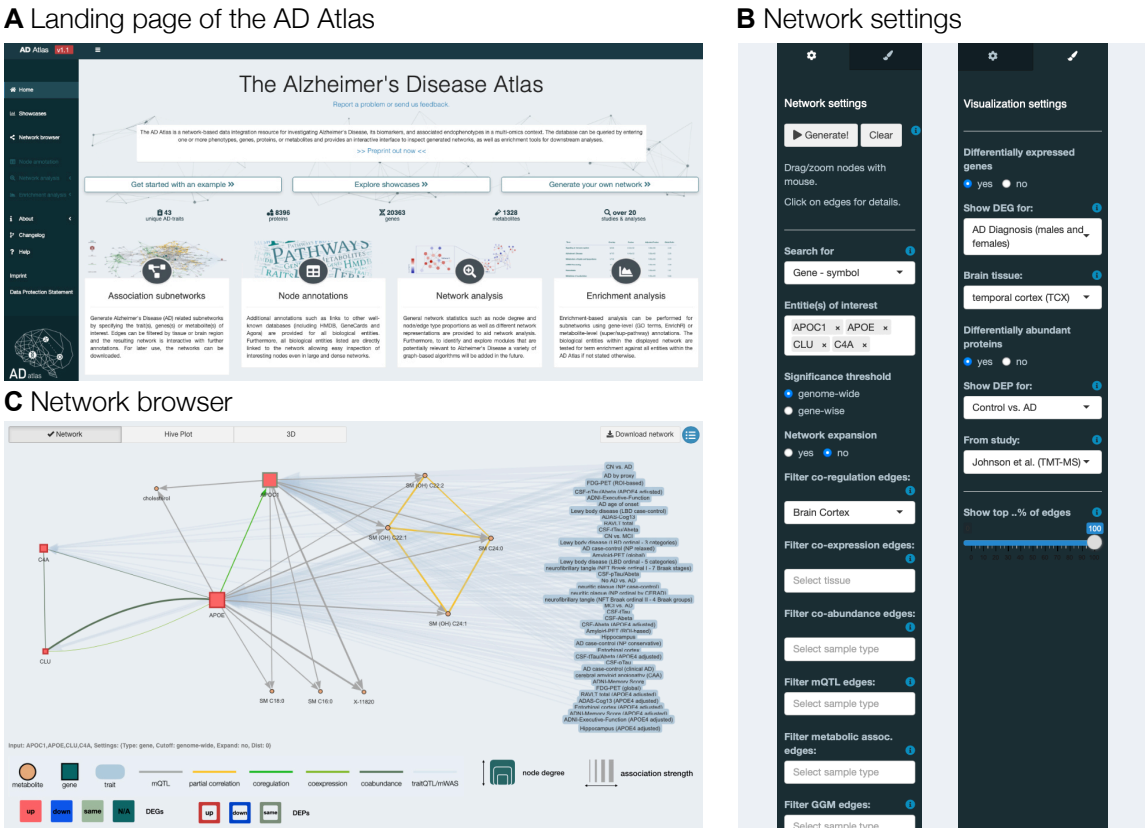


Figure 5.1: AD Atlas user interface. Users can generate context-specific molecular subnetworks and apply analytical tools via the interactive user interface (www.adatlas.org). (**A**) AD Atlas landing page. (**B**) In the network browser, users can use the right panel to specify parameters to generate context-specific molecular subnetworks (left). By clicking the paintbrush icon at the top of the panel, additional information on DEGs or DEPs in disease can be overlayed using the visualization options (right). (**C**) Network browser showcasing the multi-omics subnetwork surrounding the input genes *APOE*, *APOC1*, *C4A*, and *CLU*. (**D**) Schematic representation of the molecular subnetwork seen in C, showcasing the different node and edge types. For simplicity, not all nodes and edges are included. GWAS: genome-wide association study; mGWAS: GWAS with metabolic traits; MWAS: metabolome-wide association study; protein co-abundance: partial correlation between proteins; eQTL: expression quantitative trait locus; DEG: differentially expressed gene; DEP: differentially abundant protein.

network are represented by similar vectors or embeddings. This vector representation of the network can then be used in a number of downstream tasks, including clustering, community detection, and network visualization [367, 368]. To assess whether the underlying network structure captures biologically meaningful information, we then projected manually curated AD-relevant gene functions onto this representation of the network and tested whether functional domains are locally enriched in the embedding space. Visualization of the whole network content was then achieved by again reducing node embedding dimensions to 2-dimensional (D) vectors using Uniform Manifold Approximation and Projection (UMAP) [351].

5.2.2 Context-specific molecular subnetworks

Extending study results with (multi-)omics association data, e.g. to gain functional insights for disease-related molecular associations [146, 277] or to prioritize candidate causal genes in GWAS using overlapping eQTL or pQTL signals [212], is a commonly employed approach. Building upon this concept, the AD Atlas web interface enables the generation of context-specific molecular subnetworks by extending and interlinking entities of interest using statistical associations derived in large-scale studies of quantitative molecular data. Although correlational in nature, these networks provide a rich, evidence-based multi-omics context by integrating multiple layers of information, including eQTLs, pQTLs, mQTLs, gene co-expression and protein co-abundance, metabolic network reconstructions (based on partial correlations), as well as genetic and differential abundance associations with AD and associated biomarker profiles (schematically shown in Figure 5.1 D). These context-specific molecular subnetworks are dynamically retrieved based on user-specified parameters, allowing users to tailor networks to their research question at hand through, for example, limiting included associations to a specific brain region or pathway. The presentation in accessible network structures enables intuitive interpretation and further downstream analysis of the results.

Figure 5.2 illustrates the subnetwork generation process and highlights the core functionalities of the AD Atlas website. Users first construct a network of interest in a trait-, gene- or metabolite-centric manner (further details given in Chapter 4 Section 4.4.2) via the network settings panel in the network browser (Figure 5.1 B + C). A step-by-step schematic example of such a gene-centric network construction is given in Figure 5.2 B.

Additionally, these molecular subnetworks can be contextualized by applying edge filters (e.g. by sample type, tissue, or brain region) or adjusting the significance threshold for genetic associations (Chapter 4 Section 4.4.2). To gain further insights into the functional neighborhood of metabolites or genes, users can expand input entities before annotation to include the 1-step or 2-step neighbors using co-regulation (eQTL), transcript co-expression or protein co-abundance data for genes, or partial correlation data for metabolites. Once built, the networks can be visually inspected using the network browser (Figure 5.1 C) and

can be subjected to a number of downstream analyses within the AD Atlas. Information on differentially expressed genes (DEGs) and differentially abundant proteins (DEPs) can be overlaid onto the networks to investigate the extent and direction of dysregulation in AD using the visualization options (Figure 5.1 B). Here, users can choose the underlying association model (sex-specific or pooled analysis) and brain region. Furthermore, the entities in the generated network can be functionally characterized using gene set and pathway enrichment analysis (Chapter 4 Section 4.4.3).

In the following sections, we first provide evidence that the AD Atlas network structure recapitulates biological information on the level of functional annotations of genes. Next, we provide multiple proof-of-concept applications of the online resource, covering hypothesis-driven drug repositioning and exploratory analysis. For this, we built context-specific networks using either genes, clinical diagnosis, or metabolites as input and then assessed the resulting networks with regard to network structure, involvement in AD pathology (association with AD-related phenotypes or evidence of dysregulation at the metabolite, gene or protein level) and gene set enrichment analysis. To strengthen confidence in the relevance of the presented examples, we perform statistical evaluation of selected network statistics by calculating empirical p-values, derived from a background distribution of 1000 randomly generated networks using permutations of the same type of query input. The presented subnetworks were constructed by filtering co-regulation edges for the tissue 'Brain cortex' (only applied to COREGULATION edges) and applying a genome-wide significance cutoff to genetic associations (only applied to GENETIC_ASSOCIATION edges), if not stated otherwise. All other associations were included if they reached study-specific significance (as listed in Table 4.3) and were not filtered by sample type. To overlay differential analysis data, we used the setting 'AD Diagnosis (males and females)' and tissue 'TCX' (temporal cortex) for DEGs, and 'Control vs. AD' and 'Johnson et al. (TMT-MS)' for DEPs. All showcases can be interactively explored under www.adatlas.org/?showcases.

5.3 Global assessment of biological content of the AD Atlas

To explore the biological information content of the association-based multi-omics network on a global scale, we embedded a brain-specific network into a lower-dimensional vector space using a node embedding approach [350,366] with subsequent projection to 2D space [351] for visualization (Figure 5.3; see Chapter 4 Section 4.4.1 for further details). The embedding space reveals a clear structure (as compared to a random network embedding, Supplementary Figure 8.3 A), and overlaying gene nodes with differential expression data from AD cases vs. controls reveals distinct global patterns (Figure 5.3 A), indicating that genes whose expression is similarly affected in disease are located in proximal regions of the network. This is noteworthy, as differential expression information is not provided to the embedding algorithm. Annotating gene nodes with AD-relevant biological domains (manual curation of Gene Ontology (GO) terms into AD-relevant domains, see Chapter 4

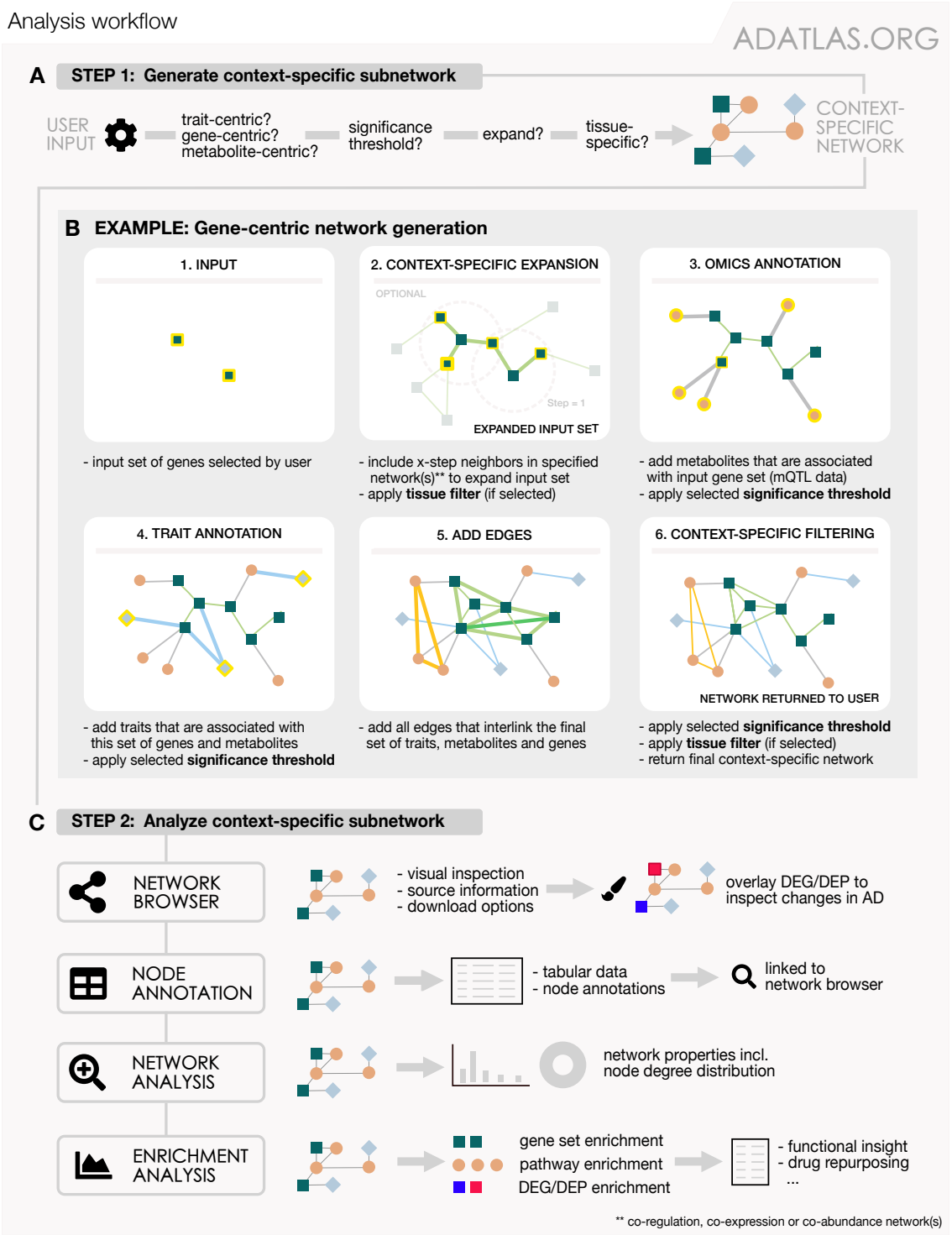


Figure 5.2: Network generation and analysis workflow. (**A**) Users can generate context-specific molecular subnetworks surrounding entities (traits, genes, metabolites) of interest. (**B**) A step-by-step description of the network generation process, here using the gene-centric approach. Statistical inter- and intra-omics links are added to the input genes provided by the user in a successive manner. Newly added nodes and edges are indicated by a yellow border and thick edges, respectively. Steps applying edge filtering (tissue-or significance-based) are indicated. (**C**) The resulting networks can be analyzed using tools, such as enrichment analysis, via our interactive user interface (`www.adatlas.org`; Figure 5.1). DEG: differentially expressed gene; DEP: differentially abundant protein.

Section 4.4.3) reveals enrichment of specific domains, i.e. biological functions, in distinct clusters obtained by hierarchical clustering of nodes in the embedding space (Figure 5.3 E). Among other examples, genes involved in immune response, a pathway previously implicated in AD pathophysiology [369], are highly enriched (p-value = 1.78×10^{-64}; one-sided Fisher's exact test) in a cluster that shows up-regulation of gene expression in disease and tight linkage (in terms of path length in the network) to AD phenotypes (Figure 5.3 B). The full set of enrichment results is provided in Supplementary Table 8.6. In conclusion, this analysis provides evidence that the association-based multi-omics network as a whole captures (AD-relevant) biological information.

5.4 Hypothesis-driven applications for drug repositioning

Little progress has been made in regard to effective and disease-modifying therapies for Alzheimer's disease [301]. Therefore, to accelerate the path to effective intervention strategies, drug repositioning - the application of available and approved compounds in a novel disease context - has gained increasing attention as a promising alternative to *de novo* drug development [307]. The following sections will highlight the potential of the AD Atlas to advance computational repositioning efforts in AD by integrative analysis of comprehensive multi-omics data.

5.4.1 Molecular subnetwork of lipid metabolism and transport identifies known repositioning candidates

In recent years, multiple genetic risk factors for late-onset Alzheimer's disease (LOAD) have been identified through GWAS [239]. The earliest was the discovery of the genetic risk exerted by the ε4 allele of *APOE* [237], which was followed by the identification of several risk variants in cerebrospinal fluid clusterin (*CLU*), also referred to as *APOJ* [370, 371]. Both proteins are involved in lipid metabolism and transport and we hypothesized that these mechanisms could potentially be targeted by available repositioning candidates. We, therefore, queried the AD Atlas using the two genes as input and expanded to the 1-step functional neighborhood defined by gene co-expression, co-regulation data, and protein co-abundance data (Chapter 4 Section 4.4.2). The resulting network provides the molecular context around these AD-associated genes by integrating multiple layers of multi-omics information (Figure 5.4 A). As expected, *APOE* displays multiple direct associations with a large number of AD-related phenotypes ranging from disease status (control vs. AD) to CSF and imaging biomarkers. *CLU* is strongly associated with disease status and AD-by-proxy [275]. The subnetwork contains 245 protein-coding genes in total, which are associated with 101 metabolites. 57 genes are genetically linked to AD (endo-)phenotypes through genome-wide significant associations. Next, we overlaid differential gene expression and protein abundance data from large-scale case-control studies of results from Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) [328,330]. This

was done using the visualization options of the AD Atlas network browser (accessible in the right panel by clicking the paintbrush icon; Figure 5.1 B) to further characterize the direction of dysregulation in disease (Figure 5.4 A). This revealed a significant up-regulation ($P_{emp} = 0.035$, Supplementary Figure 8.4) on the transcript levels of both *APOE* and *CLU* as well as their functional neighborhood in individuals with AD compared to healthy individuals (142 up-regulated, 18 down-regulated). Protein abundance data showed very similar patterns with significant overall up-regulation ($P_{emp} = 0.029$, Supplementary Figure 8.4) with 54 proteins more abundant in AD patients and 24 showing lower abundance.

As a subnetwork around two central AD risk genes are expected to show significant enrichment in terms of disease relevance, we next explored the potential of this network to identify drug repositioning candidates as another means to assess the utility of the AD Atlas. For this, we performed gene set enrichment analysis on all the genes in this subnetwork using molecular drug signatures from the EnrichR database [372] (left-hand panel under "Enrichment analysis – Gene set enrichment – enrichR"). In total, we obtained 52 unique compounds and cross-referenced this list with drugs that have been previously proposed for and tested in clinical trials [373]. Although this is a rather simplistic approach that ignores effect direction or strength, we were able to identify multiple promising candidates among the top hits, i.e. their associated lists of genes affected in drug screens overlapped most significantly with the *APOE*/*CLU* context network and they have been, or are currently, tested in clinical trials. This includes Valproate, a drug with antiepileptic properties, the anti-diabetic drug Rosiglitazone, and Fluoxetine, a selective serotonin reuptake inhibitor (SSRI) (upper table in Figure 5.4 B). However, clinical trials have failed to show a significant effect on cognition or function for both Rosiglitazone [374,375] and Valproate [376], with the latter also displaying severe toxic effects. SSRIs, especially fluoxetine, showed promising effects on AD pathology in animal models and improved cognition in a meta-analysis of short-term human trials of dementia patients with depression [377, 378]. However, further large-scale trials are needed to verify if fluoxetine offers benefits in AD patients without depression.

With the exception of Rosiglitazone, these drugs up-regulate genes in the network. However, the subnetwork surrounding *APOE* and *CLU* displays an up-regulation signature in disease and we hypothesize that drugs that perturb these genes in an opposing manner may be the most promising candidates to exert AD-relevant therapeutic effects [307]. This may also provide a partial explanation for severe side effects; if the drug and disease perturb gene expression in the same direction this could lead to aggravated symptoms or accelerated disease progression, given that the observed transcriptional changes are not a compensatory mechanism. To test this hypothesis, we repeated the enrichment analysis using a library comprised of only down-regulation signatures (EnrichR, gene set 'Drug Perturbations from GEO down'), resulting in a list of 100 unique compounds, which were

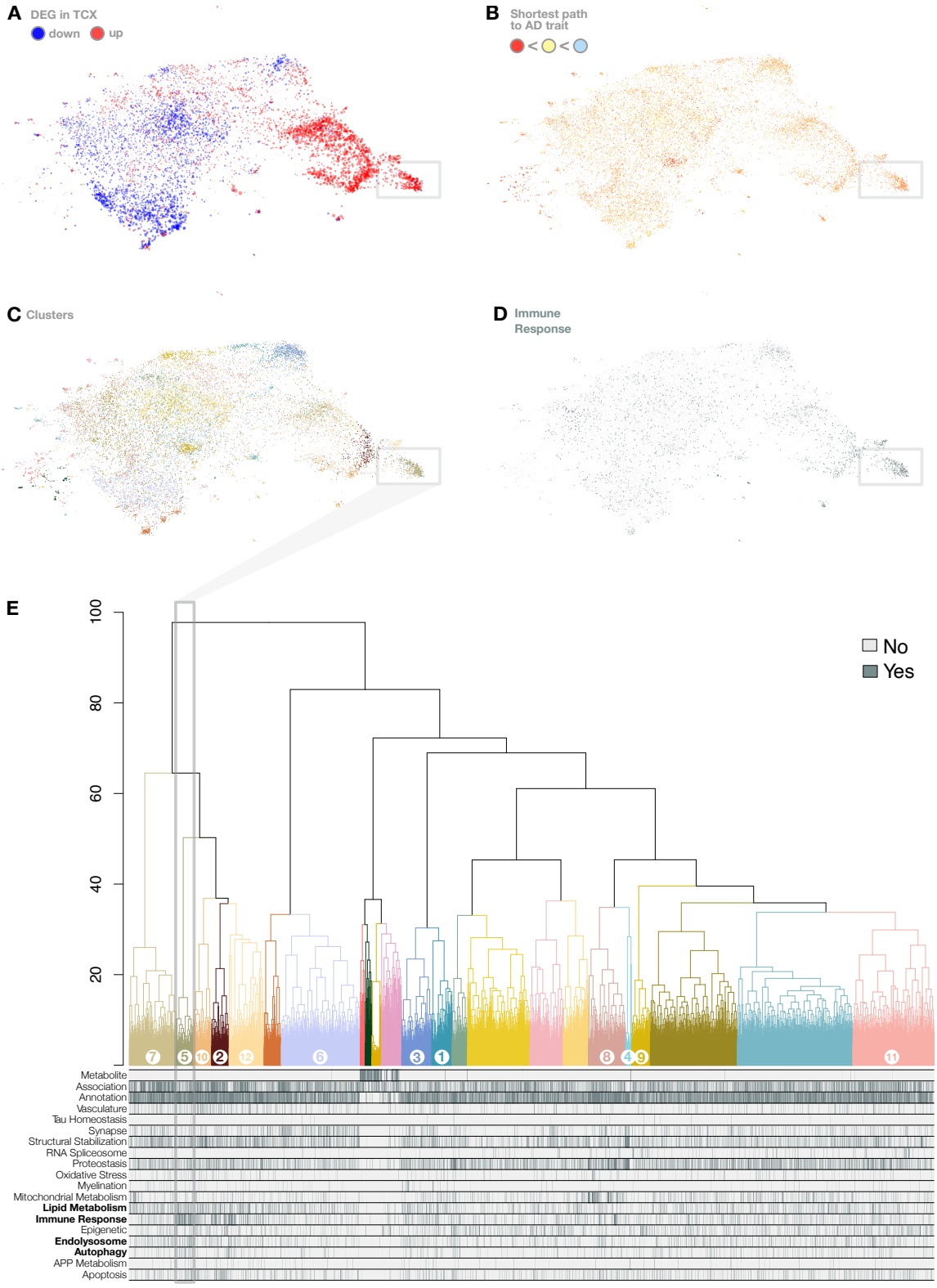

A
DEG in TCX
down
up
B
Shortest path
to AD trait
C
Clusters
D
Immune
Response
E
100
80
60
40
20
No
Yes
7
5
10
2
12
6
3
1
8
4
9
11
Metabolite
Association
Annotation
Vasculature
Tau Homeostasis
Synapse
Structural Stabilization
RNA Spliceosome
Proteostasis
Oxidative Stress
Myelination
Mitochondrial Metabolism
Lipid Metabolism
Immune Response
Epigenetic
Endolysosome
Autophagy
APP Metabolism
Apoptosis

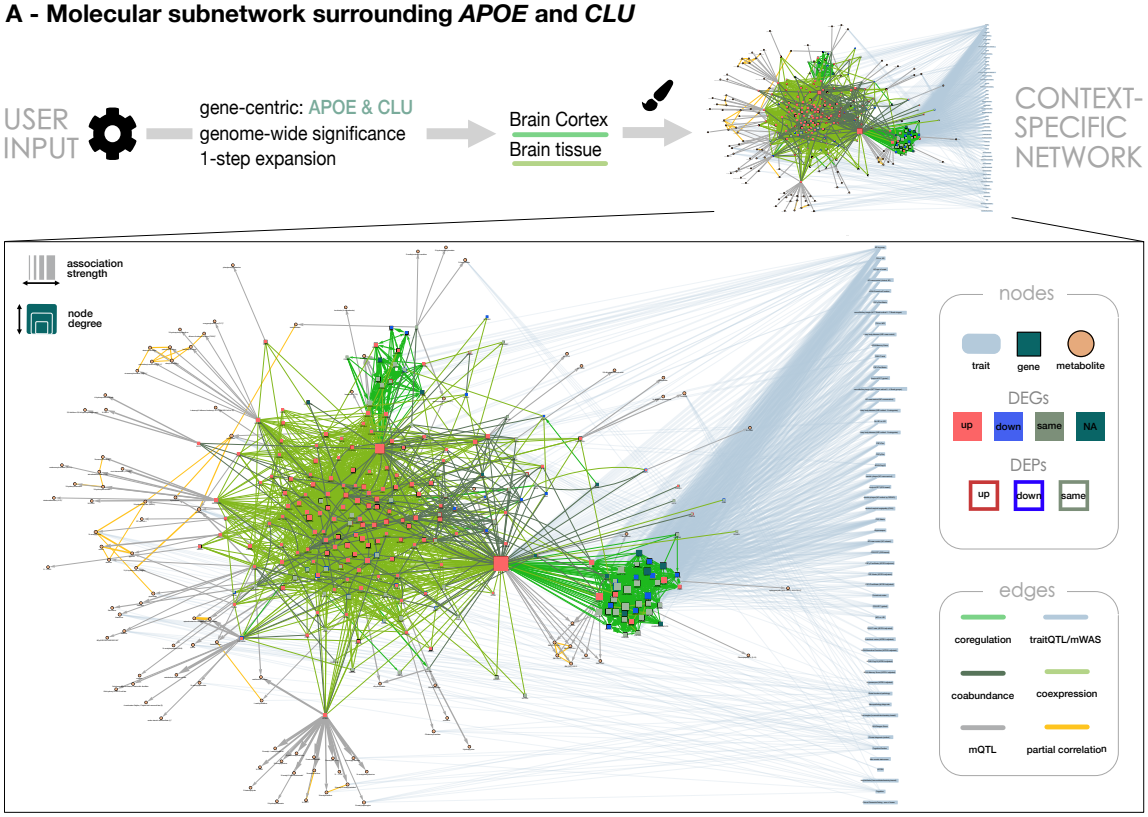
Figure 5.3: Global assessment of biological information content in the AD Atlas network structure. The nodes of the AD Atlas network were embedded into 89D vector space and projected to 2D for visualization, as described in Chapter 4. (**A**) Overlay of differential expression data from the temporal cortex (TCX) shows a global pattern of regulation. Nodes are scaled by effect size and colored by effect direction (blue: down-regulated, red: up-regulated in AD). (**B**) Shortest path of each entity (gene, metabolite) to an AD trait (red: path length 1, i.e. directly associated, orange: 2, yellow: 3 or 4, blue >4). The mean length of all shortest paths is 2.25. (**C**) Coloring according to membership of clusters determined by hierarchical clustering with cut h=30. (**D**) Immune response-related genes are colored grey. (**E**) Dendrogram of hierarchical clustering with bands below indicating whether: i) entities in clusters are metabolites, ii) have been associated with AD (through association studies or differential expression analysis), or iii) have been annotated with an AD-relevant function. This reveals the enrichment of certain biological domains in distinct regions of the visualized embedding, indicating proximity in the AD Atlas network (see Supplementary Figure 8.3 B). Significantly enriched biological domains (p-value $\leq 2.17 \times 10^{-3}$, Fisher's exact test) in cluster 5 (grey box), are emphasized in bold. Numbered clusters are significantly enriched for at least one biological domain (full results listed in Table 8.6).

again cross-referenced with clinical trials. The results of this analysis can be seen in the lower table of Figure 5.4 B.

Interestingly, the list of False Discovery Rate (FDR)-significant compounds includes Levetiracetam, a medication that is used to treat epilepsy and is currently being studied to determine whether or not it is able to improve synaptic function and reduce amyloid-induced neuronal hyperactivity as a disease-modifying therapy [379]. With Levetiracetam being investigated in multiple phase II trials (NCT02002819, NCT03489044, NCT03875638) and a low-dose formulation (AGB101) currently being tested in phase II (NCT03461861) and phase III trials (NCT03486938), Levetiracetam is one of the most represented agents among ongoing clinical trials (as of February 2020) [373]. We further investigated the genes in this subnetwork surrounding *APOE* and *CLU* that are affected by Levetiracetam. To this end, we queried the AD Atlas using the genes listed in Figure 5.4 B (via URL query parameters; Supplementary Figure 8.5 A). This analysis revealed that of the 27 genes, 20 show a dysregulation at the messenger ribonucleic acid (mRNA) level in AD with 19 in opposing direction to the drug signature (up-regulated in AD; $P_{emp} = 0$) and only one showing down-regulation. This up-regulation signature is also observed on the proteome level (14 more abundant in AD; $P_{emp} = 0.006$).

Another interesting candidate that was identified by this analysis is Candesartan, an angiotensin receptor blocker typically used for the treatment of hypertension. Of the 26 genes that are affected by the drug in the subnetwork, more than half also display a perturbed transcriptional signature in AD (13 up-regulated; $P_{emp} = 0.004$) and two are down-regulated in the TCX of AD patients (Supplementary Figure 8.5 B). Again, this up-regulation is also observed for proteins encoded by these genes (13 more abundant in AD; $P_{emp} = 0.01$). Candesartan has recently been studied in a phase II trial (NCT02646982) to investigate its effect on individuals with mild cognitive impairment that are positive for AD biomarkers. Although no results have been published for this trial, previous clinical

A - Molecular subnetwork surrounding *APOE* and *CLU*



B - Selected gene set enrichment results

Drug Perturbations from GEO 2014 (EnrichR)	*rank*	*overlap**	*q-value***	*OR****	*genes*
Rosiglitazone mus musculus gpl1261 control gds4036 chdir down	1	21/342	9.43e-07	5.68	APP, IL33, SCARA5, SDC2, TMEM176A, EPHX1, HTRA1, PCOLCE, LAMC1, C4B, CST3, PRRC1, SERP1, OLFML3, AXL, SPP1, SOX9, VIM, PTGDS, BCL2L2, DDR2
Fluoxetine mus musculus gpl1261 gds2803 chdir up	3	18/281	1.29e-06	5.88	APP, TTYH1, FAM107A, SLC1A2, STMN4, AQP4, PTN, ATP1B2, CLU, CNN3, TMEM47, CST3, GJA1, EDNRB, CPE, VIM, ITM2B, FDFT1
Rosiglitazone rattus norvegicus gpl341 adipose tissue gds3850 chdir down	4	19/318	1.29e-06	5.47	APP, TRIL, TMEM176A, HTRA1, LAMC1, NDRG2, S100B, CLU, FBLN2, DHRS3, C4B, SELENBP1, SERP1, EFEMP1, AXL, DPYSL2, TGFBI, ITM2B, DDR2
Valproic acid homo sapiens gpl6883 gse26940 chdir up	5	79/3508	1.29e-06	2.27	APP, SRPX, SLC44A3, TRIL, ZNF45, VLDLR, LAMC1, ANTXR2, PIK3C2A, GLI3, TMEM47, GJA1, SERP1, HEPH, MDK, DPYSL2, DPYSL3, ZNF404, CPNE3, SOX9, APPL2, WLS, TGIF1, SREBF1, CSNK1G3, SCARA3, SLC2A10, METRN, PHKA1, PCOLCE, SORCS2, ACOX2, PLSCR4, DDR2, FTL, SET, CCDC25, HTRA1, ITPR2, SLC1A3, AK3, PBXIP1, PRCP, PIK3R1, PTN, NDRG2, FBLN2, CST3, SELENBP1, PRRC1, GPNMB, OLFML3, FRZB, TSPAN6, SPP1, LRIG1, FYN, APOE, KCNN4, PLTP, FDFT1, ABCA1, SCD5, EPHX1, CYBRD1, PRSS35, PHKB, PC, SYT11, BCL3, SSPN, CPE, TGFBI, VIM, SFXN5, PLCD1, BCL2L2, SNTB1, ITM2C
Drug Perturbations from GEO down (EnrichR)	*rank*	*overlap**	*q-value***	*OR****	*genes*
Candesartan DB00796 rat GSE2739 sample 2687	2	26/369	3.20e-10	6.72	APP, SERPINE2, SDC4, SLC1A3, STMN4, AQP4, NDRG2, CLU, CST3, VTN, GJA1, SPOCK2, APOD, PHGDH, APOE, PTGDS, NTSR2, SREBF1, ATP1B2, S100B, AGT, BCAN, SYT11, DNAJA4, TMBIM6, ITM2B
Levetiracetam DB01202 rat GSE2880 sample 2777	17	13/181	1.99e-05	6.53	WFS1, STMN4, ENO1, PTN, NDRG2, S100B, CST3, DPYSL2, TRIM35, CPE, APOE, ITM2B, ITM2C
Levetiracetam 5284583 rat GSE2880 sample 2669	22	19/399	2.21e-05	4.29	APP, ENO1, ATP1B2, CLU, AGT, CNN3, GCSH, GLUD1, EDNRB, TRIM35, PHGDH, ACSBG1, TMBIM6, APOE, VIM, PTGDS, ITM2B, NTSR2, ITM2C

*overlap network (nominator) with drug signature gene set (denominator) **adjusted *p*-value using Benjamini-Hochberg ***odds-ratio

Figure 5.4: *APOE/CLU* subnetwork identifies repositioning candidates. (**A**) Multi-omics subnetwork surrounding *APOE* and *CLU* as contained in the AD Atlas. (**B**) Gene set enrichment analysis for drug-associated gene expression changes using EnrichR reveals previously proposed candidates Valproate, Fluoxetine, and Rosiglitazone among the drugs most significantly affecting the subnetwork (upper table). When focusing on signatures that are opposed to the overall change in AD, we identify Levetiracetam and Candesartan as most promising (lower table). Genes that overlap between drug signature sets (column 'genes') and the network are color-coded to indicate the direction of expression change or protein abundance in the disease (red: up-regulation, blue: down-regulation). Genes with opposing directions of change at the transcript (color code) and protein level are underlined.

trials suggest beneficial neurocognitive effects following Candesartan treatment in older individuals with hypertension and mild cognitive impairment [380, 381]. Of note, investigation of the network structure in the AD Atlas revealed that genes down-regulated by Levetiracetam and Candesartan lie within distinct genetic loci and are connected at the mRNA and protein level (Supplementary Figure 8.5). Furthermore, both affect the AD-risk genes *APOE* and *CLU*, which were provided as input, as well as amyloid beta precursor protein (*APP*).

In summary, we showcase how the AD Atlas can provide additional AD-relevant insights for known AD risk genes. By applying a simple enrichment approach requiring minimal analysis steps, we identify multiple drug repositioning candidates that are either being tested or have been tested in clinical trials, providing evidence for the relevance of the generated results. Through additional analysis using the AD Atlas we propose Levetiracetam and Candesartan as the most promising candidates as their associated list of genes affected in drug screens overlaps significantly with the molecular context network surrounding *APOE* and *CLU* and they affect disease-perturbed genes in an opposing manner.

5.4.2 Disease-associated molecular subnetwork provides a global view on AD

The previous analysis uses a user-defined set of AD-related genes as starting point for the repurposing analysis, allowing an in-depth and focused view of specific aspects (lipid metabolism) of the disease. To move beyond that, the AD Atlas also provides a trait-specific entry-point that enables the generation of global, data-driven views on AD and its associated (endo-)phenotypes, enabling more comprehensive analyses. Here, entities (genes and metabolites) associated with a specific (set of) AD-related trait(s) identified in large-scale GWAS and MWAS, are embedded into their multi-omics context that is annotated with metabolic and genetic associations as well as intra-omics links. Using this entry point, we repeated the analysis steps outlined in the previous example. We selected the results of large-scale AD case-control studies from International Genomics of Alzheimer's Project (IGAP) [272, 278] and Alzheimer's Disease Neuroimaging Initiative (ADNI) (trait 'CN vs. AD') as input to build a global molecular subnetwork of genetic risk for AD. The network can be seen in Figure 5.5 A and consists of 380 genes and 93 metabolites. The network shows multiple tightly connected clusters of genes, which are indicative of the underlying genetic architecture (genetic loci are tightly connected through co-regulation edges). This includes, for example, the *CR1* (chromosome 1), *BIN1* (chromosome 2), *ZCWPW1/NYAP1/PILRA* (chromosome 7), *PICALM* (chromosome 11) and *APOE* (chromosome 19) locus. To investigate the degree and direction of perturbation at the mRNA and protein level, we used the AD Atlas visualization options to overlay results of differential gene and protein analysis. This revealed both up-regulation (70 transcript-level, 26 protein-level) and down-regulation (52 transcript-level, 31 protein-level) of the

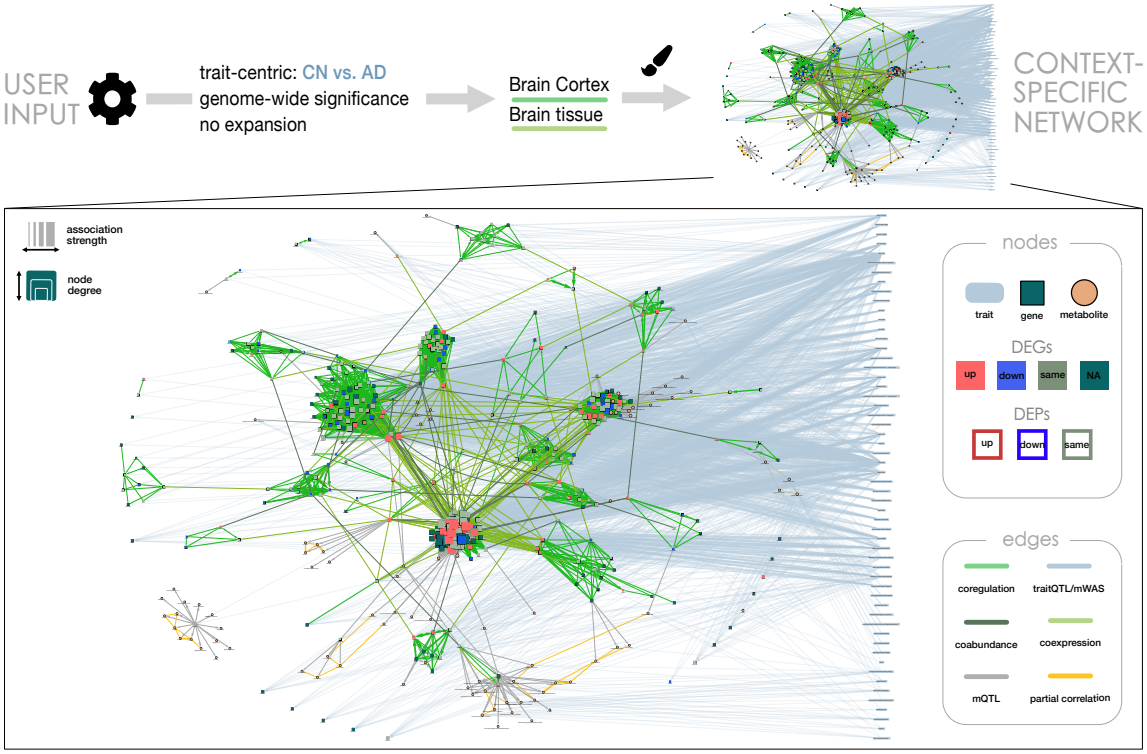
AD-related subnetwork which is, however, not statistically significant (Supplementary Figure 8.6).

Utilizing the previously described concept of gene set enrichment to identify repositioning candidates, we perform the analysis using the drug perturbations library from GEO via EnrichR. The only two drugs that display a significant overlap of genes with the trait-centric network after multiple testing correction are Letrozole and Rosiglitazone (upper table in Figure 5.5 B). Unfortunately, both of these agents do not seem to be promising candidates. Letrozole is an aromatase inhibitor used to treat estrogen receptor–positive breast cancer in postmenopausal women and aromatase inhibitors have been associated with adverse effects including negative effects on cognition and potential long-term neural effects [382]. Furthermore, Rosiglitazone, as discussed previously, has failed to show significant effects on cognition in clinical trials [374,375]. Therefore, we focused the enrichment analysis again on drug-induced signatures of down-regulated gene expression (gene set 'Drug Perturbations from GEO down' from EnrichR).

This analysis yielded multiple significant hits (in total eight unique compounds), which we again screened for compounds in clinical trials to highlight the potential of such networks to identify AD-relevant drug candidates (lower table in Figure 5.5 B) and capture biologically relevant aspects of the disease. One promising finding was Citalopram, a SSRI used in the treatment of depression. Using the AD Atlas, we further investigated whether the genes perturbed by this agent are also perturbed in AD and if its effects on transcript levels are in an opposing direction. Citalopram affects a total of 16 genes in the trait-centric subnetwork and more than half are differentially expressed in AD (nine up-regulated, $P_{emp} = 0.008$; and one down-regulated $P_{emp} > 0.05$; Supplementary Figure 8.7 A). Furthermore, Escitalopram, the (S)-stereoisomer of Citalopram, is currently being studied as a treatment for agitation in AD patients in a phase III trial (NCT03108846) and has also entered a phase I trial as a cognitive enhancer (NCT03274817).

Another interesting finding was Etanercept, a tumor necrosis factor α (TNFα) inhibitor that is used to treat autoimmune diseases. Elevated levels of TNFα, an inflammatory cytokine, in the brain have been linked to AD and proposed as a potential therapeutic target [383]. Repeating the analysis steps outlined for Citalopram, we find that the Etanercept gene set has 14 overlapping genes of which five show up-regulation in AD ($P_{emp} = 0.03$; Supplementary Figure 8.7 B). It is interesting to note that the subnetwork also shows a significant overlap with genes that are up-regulated by Etanercept (gene set 'Drug Perturbations from GEO up' from EnrichR, $\mathrm{P}_{FDR} = 5.88e^{-03}$), indicating that the drug may target both compensatory and disease mechanisms. Furthermore, there is some overlap between the up- and down-perturbation gene sets which may point to a degree of variability within the response. Safety and tolerability of Etanercept in AD has been established in a small randomized, placebo-controlled, double-blind phase II trial (NCT01068353) [384] but the drug has not been studied in a phase III trial, despite multiple large observational

A - Multi-omics contextualization of large-scale AD case-control GWAS and MWAS



B - Selected gene set enrichment results

Drug Perturbations from GEO 2014 (EnrichR)	*rank*	*overlap**	*q-value***	*OR****	*genes*
Letrozole homo sapiens gpl3921 gse33366 chdir down	1	22/404	5.74e-03	3.10	DDR1, EPHX2, HLA-B, B3GALT4, HLA-C, TAP1, IFI35, CLU, HLA-E, PSMB9, C4B, C4A, AZGP1, BCAM, BCL3, FLOT1, TIMP3, LY6G5C, CFB, HLA-DRB1, CD55, ZKSCAN1
Rosiglitazone homo sapiens gpl570 gds2453 chdir down	2	21/403	8.72e-03	2.95	VAT1, CR1, GPX4, MS4A7, HLA-B, HLA-C, NR1H3, TREM2, DTX4, AIF1, MS4A4A, MS4A6A, SDCBP, HLA-DRA, APOE, CD46, CD68, CLIC1, HLA-DRB1, HSPA1B, PICALM
Drug Perturbations from GEO down (EnrichR)	*rank*	*overlap**	*q-value***	*OR****	*genes*
Etanercept DB00005 human GSE36177 sample 2596	5	14/223	1.42e-02	3.56	HLA-DRB5, MS4A7, KIR3DL1, AOAH, KIR2DL1, PRRC2A, MS4A6A, HLA-DMA, HLA-DMB, HLA-DRA, UGGT1, HLA-DRB1, CD55, HSPA1B
Citalopram DB00215 human GSE7036 sample 3292	7	16/288	1.67e-02	3.14	HLA-DRB5, PVRIG, HLA-B, HLA-C, IFI35, PSMB8, HLA-E, HLA-DMA, HLA-DMB, STAG3, HLA-DRA, HLA-DQA1, HLA-DRB1, HSPA1B, HLA-DPA1, HLA-DQB1

*overlap network (nominator) with drug signature gene set (denominator) **adjusted *p*-value using Benjamini-Hochberg ***odds-ratio

Figure 5.5: AD case-control subnetwork identifies potential repositioning candidates. (**A**) Multi-omics contextualization of large-scale AD case-control GWAS and MWAS studies. The network was built using the trait 'CN vs. AD' as input for trait-centric subnetwork generation. (**B**) Gene set enrichment analysis for drug-associated gene expression changes in drug perturbation gene sets from Gene Expression Omnibus (GEO) via EnrichR reveals Letrozole and Rosiglitazone as the only drugs significantly affecting the subnetwork. By focusing on down-regulation signatures, we obtain more significant hits, among which Etanercept and Citalopram seem to be the most promising candidates. Genes that overlap between drug signature sets (column 'genes') and the network are color-coded to indicate the direction of expression change or protein abundance in the disease (red: up-regulation, blue: down-regulation). Genes with opposing directions of change at the transcript (color code) and protein level are underlined.

studies indicating a reduced risk of AD among patients treated with Etanercept [385]. Further supporting evidence is provided by a second generation TNFα inhibitor, selective for the soluble form of TNFα, that has also shown promising results in preclinical studies [386, 387] and was recently investigated in a Phase I trial (NCT03943264). Both Etanercept and Citalopram seem to affect the immunoregulatory human leukocyte antigen (HLA) complex, which has been implicated in neurodegenerative diseases, including AD [278, 388].

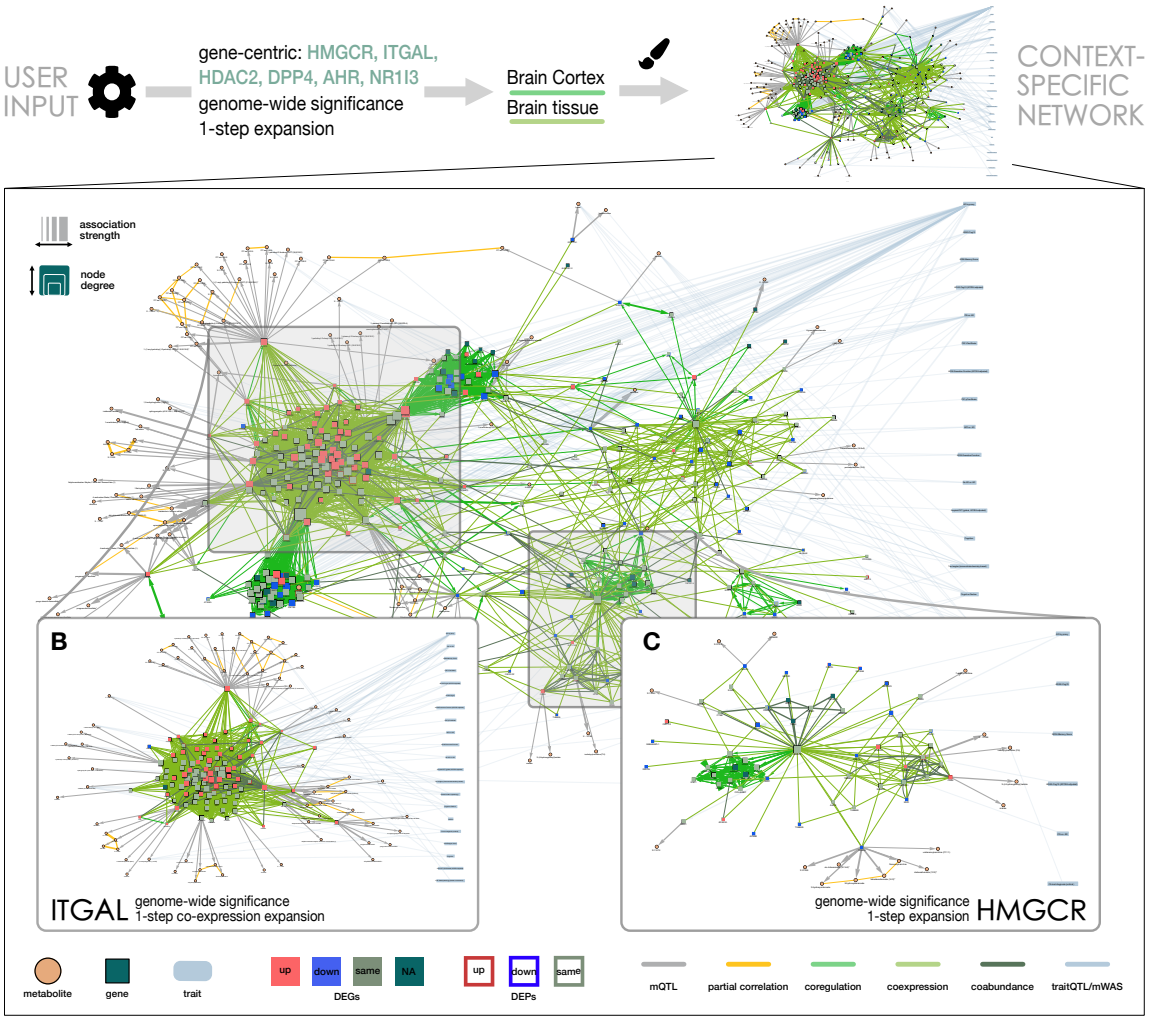
In conclusion, the AD Atlas enables the generation of global, disease-related molecular networks without the need to perform complex data analysis. These networks can be used in downstream analysis within the web-based user interface, for instance, to identify plausible repositioning candidates using gene set enrichment. In addition, the networks and lists of contained entities can be directly downloaded for offline follow-up analyses. Despite coming from diverse sources and cohorts, the totality of data collected in the AD Atlas can provide valuable insights into the underlying mechanisms of disease, as we demonstrate with the identification of Etanercept and Citalopram as promising repositioning candidates through potential modulation of neuroinflammatory pathways.

5.4.3 Statin target *ITGAL* links to neuroinflammation through TREM2 signaling

Statins are a class of lipid-lowering drugs that are used to reduce the risk of cardiovascular diseases, such as atherosclerosis and peripheral artery disease. Statins exert their primary therapeutic effect by inhibiting the rate-limiting enzyme, 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCR), in the cholesterol-producing mevalonate pathway. In addition, statins have been associated with a wide range of secondary effects [389, 390]. Observational studies have reported a possible association between statin use and reduced risk of AD [391, 392, 393], although the evidence has been inconsistent with reported differences between patient subpopulations as well as individual statin drugs [394, 395, 396]. To investigate this potential link to AD pathophysiology we used the drug targets of statins, as annotated in DrugBank [397] (Supplementary Table 8.7) and constructed their molecular context network in the AD Atlas (Figure 5.6 A). To this end, we used *HMGCR*, *ITGAL*, *HDAC2*, *DPP4*, *AHR* and *NR1I3* genes as input, added their 1-step co-expression, co-regulation and protein co-abundance neighbors by network expansion and colored genes according to their differential expression in AD patients using the visualization options of the AD Atlas. Of the 321 genes contained in the resulting network, 65 genes are genetically linked to AD (endo-)phenotypes ($P_{emp} = 0.044$; Supplementary Figure 8.8) and about one-third (n=126) show differential expression at the transcript level in the TCX of AD patients ($P_{emp} > 0.05$; Supplementary Figure 8.8).

Visual inspection of the network indicates two tightly connected clusters surrounding *ITGAL*, one driven by co-regulated genes (vibrant green edges) and one driven by co-

A - Molecular subnetwork surrounding statin targets annotated in DrugBank



D - Top 5 gene set enrichment results

WikiPathway 2021 human (EnrichR)	*overlap**	*q-value** (ITGAL***)*	*OR*****	*genes*
TYROBP causal network in microglia WP3945	17/61	**1.92e-14** (2.08e-23)	24.95	RBM47, IGSF6, PLEK, ITGB2, TCIRG1, SLC1A5, TGFBR1, PYCARD, APBB1IP, CD4, SLC7A7, ADAP2, BIN2, ELF4, CD37, GAL3ST4, TMEM106A
Cholesterol Biosynthesis Pathway WP197	7/15	**2.46e-07** (-)	54.82	FDPS, SQLE, HMGCS1, SC5D, MSMO1, MVD, HMGCR
Cholesterol metabolism (includes both Bloch and Kandutsch-Russell pathways) WP4718	10/46	**2.67e-07** (-)	17.54	FDPS, SQLE, EBP, HMGCS1, FASN, SC5D, MSMO1, MVD, HMGCR, SREBF2
Microglia Pathogen Phagocytosis Pathway WP3937	9/40	**9.26e-07** (1.86e-09)	18.28	HCK, FCER1G, ARPC1B, ITGB2, RAC2, TREM2, PTPN6, PIK3CG, VAV1
Mevalonate pathway WP3963	5/7	**1.40e-06** (-)	155.67	FDPS, HMGCS1, MVD, HMGCR, ACAT2

*overlap network (nominator) with pathway gene set (denominator) **adjusted *p*-value using Benjamini-Hochberg
***enrichment for 1-step co-expression network surrounding ITGAL ****odds-ratio

Figure 5.6: Subnetwork of statin targets link to TYROBP signaling. (**A**) Multi-omics subnetwork surrounding the statin targets annotated in DrugBank; *HMGCR*, *ITGAL*, *HDAC2*, *DPP4*, *AHR* and *NR1I3*, as well as (**B**) a co-expression subnetwork surrounding only *ITGAL* and (**C**) a co-expression, co-regulation and co-abundance subnetwork surrounding only *HMGCR*. (**D**) Gene set enrichment analysis using EnrichR (gene set 'WikiPathway'). Genes that overlap between pathway sets (column 'genes') and the network are color-coded to indicate the direction of expression change or protein abundance in disease (red: up-regulation, blue: down-regulation).

expressed genes (light green edges). Another co-regulation cluster is formed by co-regulated genes surrounding *NR1I3*, of which *ARHGAP30*, *FCGR2A*, *FCGR3A* and *FCER1G* are tightly linked with the *ITGAL* co-expression module. Both co-regulation clusters contain multiple genes that are dysregulated in AD. The direct neighborhood of each of the six input genes reveals that all targets show a dysregulation at the transcript or protein level in AD, either through direct evidence or via their neighbors. For example, while the primary statin target *HMGCR* (Figure 5.6 C) does not show evidence for differential regulation or association to AD phenotypes, 37% of its direct neighbors in the network (20 out of 54 genes) show either up- (n=5) or down- (n=15) regulation at the transcript level and ten (four up- and six down-regulated) at the protein level. However, this does not significantly deviate from the random background distribution (Supplementary Figure 8.9), suggesting that a potential benefit from statin treatment may not be directly conveyed through the modulation of the cholesterol synthesis pathway.

To functionally characterize the pathways targeted by statins further, we used the gene set enrichment functionality of the AD Atlas. Here, we found significant enrichment for the TYROBP causal network (EnrichR, gene set 'WikiPathways', $P_{FDR} = 1.92e^{-14}$), an immune- and microglia-specific module that has been implicated in LOAD [285]. This enrichment is driven by the co-expression network surrounding *ITGAL* (Figure 5.6 B), also known as *CD11a*, a subunit of the integrin leukocyte function-associated antigen-1 (LFA-1) which is involved in a variety of immune-related functions [398]. Interestingly, only a subset of statins (Rosuvastatin, Lovastatin, Simvastatin, Pitavastatin) target *ITGAL* which may explain the heterogenous results reported in studies. Furthermore, the co-expression neighborhood of *ITGAL* shows substantial dysregulation in disease with 50 of 103 genes differentially expressed in AD (49 up-regulated, one down-regulated; $P_{emp} > 0.05$), 13 genes encoding for proteins that are up-regulated in disease ($P_{emp} = 0.044$) and 16 genes genetically associated with AD ($P_{emp} = 0.059$; Supplementary Figure 8.10). TYROBP (Dap12) is an adaptor molecule involved in the transduction pathway of *TREM2* as well as *CD33*, a known AD risk gene [338], and CR3 (*ITGAM* and *ITGB2*). Both *TREM2* and *ITGB2* are up-regulated in disease and contained in the network shown in Figure 5.6 B.

Given those results, an alternative hypothesis on the controversial effects of statins on AD risk can be formulated, as the identified functional link to TYROBP signaling indicates that the subset of statins targeting *ITGAL* may exert potential protective effects [399,400] through modulation of neuroinflammatory pathways [401].

5.5 Application examples for exploratory analysis

High failure rates of AD drugs in clinical trials have emphasized our incomplete understanding of crucial biological aspects of this complex disease [301]. Basic research investigating the underlying molecular disease mechanisms and mapping the trajectory of biochemical

changes in AD, will therefore be crucial to identify novel therapeutic targets and guide future drug development efforts. In the following examples, we showcase how the AD Atlas can be used to investigate and formulate new AD-related hypothesis. We focus on the contextualization of metabolic and immune-related AD findings, although there are many more areas of application.

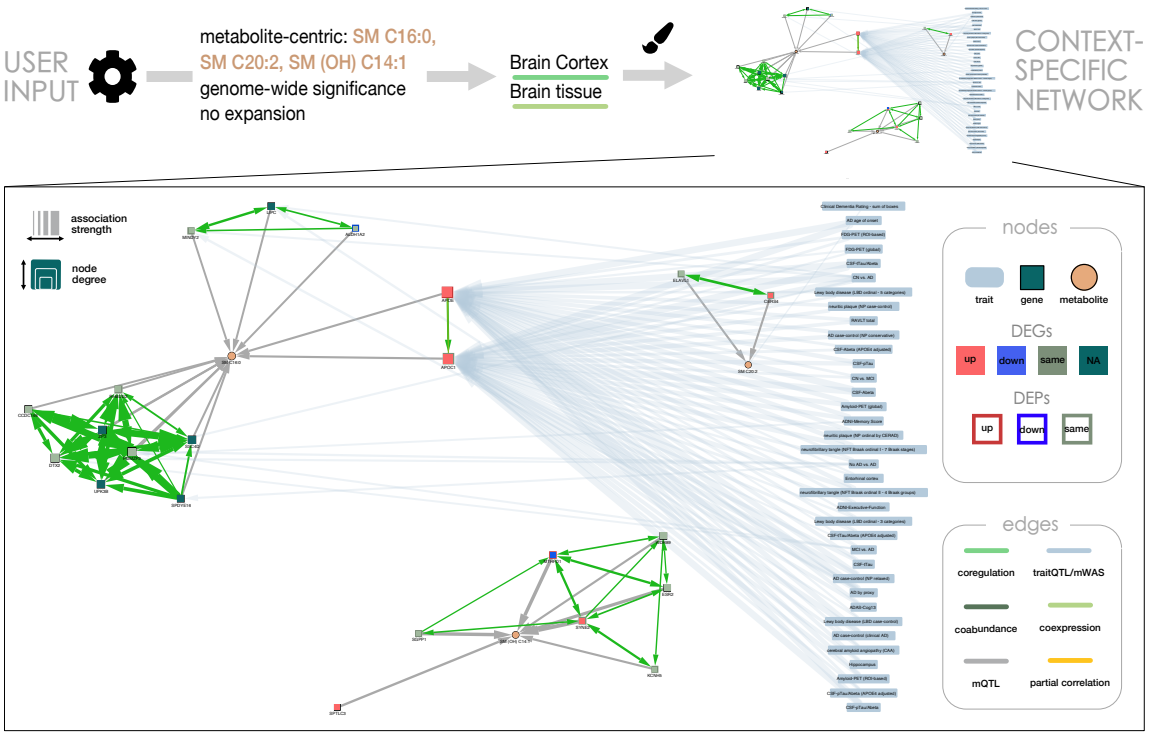
5.5.1 Contextualization of links between the sphingomyelin pathway and AD pathology

In a previous study, we identified sphingomyelin (SM) species of differing lengths to be implicated in early vs. late stages of AD [19]. More precisely, we found SM C16:0 to be associated with CSF $A\beta_{42}$ pathology, while SMs with longer fatty acid chains ($\geq$ C20) were correlated with brain atrophy and cognitive decline. This study identified three SMs associated with AD, labeled as SM (OH) C14:1, SM C16:0, SM C20:2 (of note, these vendor-specific labels, i.e. biochemical names assigned by the respective metabolomics platform provider, will be updated to better reflect the currently accepted notations for SM in future releases). We generated a metabolite-centric subnetwork via the AD Atlas user interface to contextualize these findings and gain a better understanding of their potential functional role in AD.

The generated subnetwork is shown in Figure 5.7 A. All SMs are associated with at least two genes at a genome-wide significance level. Interestingly, these particular SMs are not directly linked to each other through partial correlations. The associated genes are interlinked by co-regulation edges pointing to the same genetic loci and the individual 'SM-gene-clusters' are not connected with each other. These aspects may point to their involvement in distinct pathways or pathway steps, although the absence of an edge should be interpreted with caution, as this can also be due to lack of coverage in or insufficient power of the available studies. Furthermore, 35 of the 36 AD-related traits are associated with SM C16:0 related genes, which include AD risk genes *APOE* and *APOC1*. The SM (OH) C14:1 cluster shows no direct or indirect trait associations, while SM C20:2 is directly associated with the clinical dementia rating - sum of boxes score. Overlay of differential gene and protein expression data using the visualization options provided by the AD Atlas reveals an up-regulation at the mRNA level of 5 of 22 genes (*APOE* locus (*APOE*, *APOC1*), *SPTLC3*, *SYNE2* and *CERS4*; $P_{emp} > 0.05$; Supplementary Figure 5.7) and down-regulation of one gene (*MTHFD1*), which displays an opposing effect at the protein level (up-regulation).

To further characterize the potential functional involvement of these SM-associated genes in AD, we performed an enrichment analysis using the Reactome 2016 library via EnrichR. This identifies three genes (*CERS4*, *SPTLC3* and *SGPP1*) involved in SM *de novo* biosynthesis ($P_{FDR} = 2.53e^{-04}$, Figure 5.7 B). Interestingly, these genes have previously been identified in a study that involved multiple, time-intensive manual mapping steps [277].

A - Molecular subnetwork surrounding SM (OH) C14:1, SM C16:0, SM C20:2



B - Top 5 gene set enrichment results

Reactome 2016 (EnrichR)	*overlap**	*q-value***	*OR****	*genes*
Sphingolipid de novo biosynthesis Homo sapiens R-HSA-1660661	3/33	2.53e-04	104.99	CERS4, SPTLC3, SGPP1
Sphingolipid metabolism Homo sapiens R-HSA-428157	3/74	1.46e-03	44.27	CERS4, SPTLC3, SGPP1
Chylomicron-mediated lipid transport Homo sapiens R-HSA-174800	2/17	2.13e-03	133.09	LIPC, APOE
Lipoprotein metabolism Homo sapiens R-HSA-174824	2/34	5.20e-03	62.33	LIPC, APOE
Metabolism of lipids and lipoproteins Homo sapiens R-HSA-556833	5/659	5.20e-03	8.69	CERS4, LIPC, SPTLC3, APOE, SGPP1

*overlap network (nominator) with pathway gene set (denominator) **adjusted *p*-value using Benjamini-Hochberg ***odds-ratio

Figure 5.7: Contextualization of metabolomics-guided insights points to SM *de novo* biosynthesis. (**A**) Multi-omics subnetwork surrounding three SM species (SM (OH) C14:1, SM C16:0, SM C20:2) that are altered in biomarker defined stages of AD [19]. (**B**) Gene set enrichment analysis using the Reactome gene set accessible via EnrichR identifies three genes (*CERS4*, *SPTLC3* and *SGPP1*) involved in SM *de novo* biosynthesis. Genes that overlap between pathway sets (column genes) and networks are color-coded to indicate the direction of expression change or protein abundance in disease (red: up-regulation, blue: down-regulation).

Here, the genes were categorized into two functional categories: global sphingomyelin synthesis (*SPTLC3*, *CERS4*), and synthesis and degradation of sphingosine-1-phosphate (*SGPP1*), highlighting a possible role for sphingosine-1-phosphate and its receptors in AD pathogenesis. AD mouse models indicate a potential benefit of Fingolimod, an Food and Drug Administration (FDA)-approved sphingosine-1-phosphate (S1P) analog used for the treatment of multiple sclerosis [402, 403, 404]. Furthermore, long-term Fingolimod treatment in multiple sclerosis patients has shown positive effects on cognition [405]. Therefore,

Baloni et al. applied a drug repositioning approach by treating APP/PS1 mice with Fingolimod, finding that prolonged S1P pathway modulation can rescue both the proposed cellular mechanism of hippocampus-related memory and cognitive deficits in these mice, further supporting this pathway as a high priority target for AD [277].

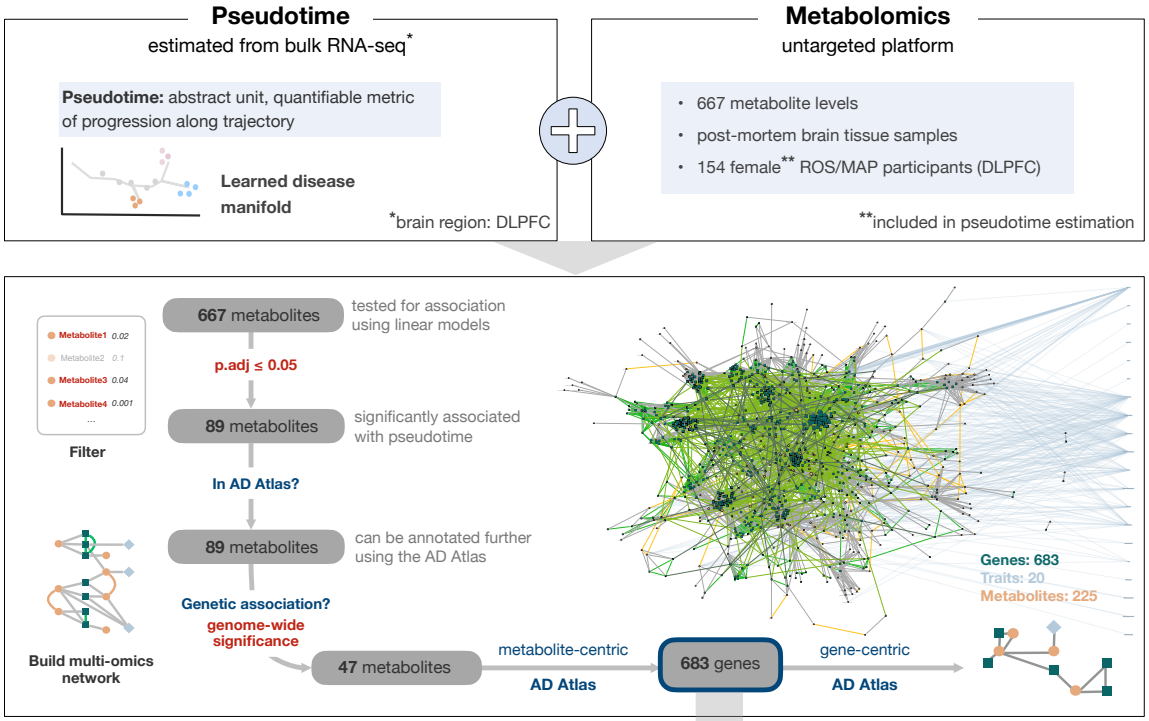
In conclusion, this analysis highlights the ability of the AD Atlas to contextualize hypotheses or findings from previous studies and thereby point to novel, AD-related insights without the need for bioinformatics resources or time-intensive manual analyses. Here, we used three sphingomyelin species that have previously been associated with AD endophenotypes [19], to further investigate their involvement in the mechanisms of disease. We find a link to the S1P pathway, replicating results of a recent study [277].

5.5.2 Multi-omics characterization of brain-based pseudotime estimates for AD progression

The cascade of molecular changes that leads to AD onset and progression remains incompletely understood. Due to the fact that post-mortem brain samples are always cross-sectional, some kind of temporal metric is needed to study disease progression. Recently, researchers have utilized manifold learning techniques to construct pseudo-temporal models of the disease using brain transcriptomics data [406]. These approaches order samples on a tree-like molecular progression trajectory according to their similarities in expression profiles and enable the estimation of so-called pseudotime – an abstract metric that quantifies the progression of an individual along this trajectory [406]. Downstream analyses provided insights into potentially disease-driving pathways, including links to mitochondrial dysfunction. Interestingly, pseudotime inferred for females only was more strongly correlated to disease status and neuropathological measures than for males only [406]. To enable a more comprehensive molecular characterization of disease progression we embedded these models into a multi-omics context using the AD Atlas. The analysis was performed in two steps and we restricted our analyses to female samples due to the aforementioned sex differences. First, we determined metabolites significantly associated with pseudotime, and second, we used these metabolites as proxies for transcriptome-derived pseudotime to build a multi-omics context network and further characterize the underlying molecular mechanisms. A schematic overview of this analysis is given in Figure 5.8 A.

Using non-targeted metabolite (n=667) measurements from post-mortem brain tissue samples (dorsolateral prefrontal cortex (DLPFC)) taken from 154 female Religious Orders Study (ROS)/Memory and Aging Project (MAP) participants [263], we tested for pairwise associations between metabolite levels and pseudotime estimates using linear regression models. In total, 89 of the 667 metabolites showed a significant association with pseudotime at $\alpha = 0.05$ after Bonferroni correction, i.e. the concentration of these molecules either significantly increased or decreased with progressing pseudotime. These significantly associated metabolites span multiple pathway classes, including amino acids (36/89), lipids

A - Multi-omics characterization of brain-based pseudotime estimates for AD progression



B - Top 6 gene set enrichment results

Reactome 2016 (EnrichR)	*overlap**	*q-value***	*OR****
Metabolism of amino acids and derivatives R-HSA-71291	32/335	**8.12e-05**	3.08
Metabolism R-HSA-1430728	108/1908	**8.12e-05**	1.83
Amino acid transport across the plasma membrane R-HSA-352230	9/31	**1.91e-04**	11.71
SLC-mediated transmembrane transport R-HSA-425407	25/268	**1.36e-03**	2.98
Histidine, lysine, phenylalanine, tyrosine, proline and tryptophan catabolism R-HSA-6788656	9/41	**1.47e-03**	8.05
Na+/Cl- dependent neurotransmitter transporters R-HSA-442660	6/19	**4.47e-03**	13.16

*overlap network (nominator) with pathway gene set (denominator) **adjusted *p*-value using Benjamini-Hochberg ***odds-ratio

Figure 5.8: Multi-omics characterization of brain-based pseudotime estimates. (**A**) Metabolites [263] significantly associated with brain-based pseudotime estimates [406] were used as pseudotime proxies in a two-step workflow to build multi-omics context networks using the AD Atlas. (**B**) Gene set enrichment analysis using the Reactome gene set accessible via EnrichR points to neurotransmission, bioenergetic, and transport pathways.

(17/89), nucleotides (10/89), and various unknown compounds (15/89). Interestingly, 84% (75/89) of these metabolites have been associated with AD-related traits in a previous brain-based study [263].

Of the 89 metabolites, 47 metabolites could be linked to genes through mQTL data using the AD Atlas (metabolite-centric network, genome-wide significance). This resulted in a total of 683 genes. Next, we extended this set of genes to include additional metabolite and trait associations as well as inter-omics links by again building a multi-omics network using the AD Atlas (gene-centric network, genome-wide significance). The resulting network

(Figure 5.8 A) was comprised of 683 genes, 225 metabolites, and links to 20 AD-related traits, including CSF amyloid pathology, brain glucose uptake measured by FDG-positron emission tomography (PET), and cognitive measures. 81% of the pseudotime associated metabolites with genetic associations (38/47) and nearly one-third of the genes contained in the network showed significant associations with AD, with transcriptional changes most pronounced in the TCX (112 genes up- and 108 genes down-regulated in AD). Lastly, we applied pathway enrichment analysis using the functionalities of the AD Atlas to derive over-represented biological processes potentially involved in disease progression. This analysis revealed functional links to neurotransmission, bioenergetics, and transport (see Figure 5.8 B), pathways previously implicated in AD pathogenesis.

In summary, by using metabolites as proxies for transcriptome-derived pseudotime, we were able to investigate the molecular underpinnings of AD progression in a multi-omics context. Our analysis provides further molecular evidence for pathways implicated in AD and emphasizes the potential of such metabolomics-guided analysis for future studies.

5.5.3 Subnetworks surrounding marker genes for homeostatic microglia and disease-associated microglia suggests a possible involvement of blood androgens

Genomic analyses in AD and animal models of the disease identified a specific activation program that drives the transition from homeostatic microglia to disease-associated microglia [289]. However, the molecular mechanisms underlying this transition remain incompletely understood and the extent to which this process involves AD susceptibility genes has not been assessed in an integrated fashion. We used the AD Atlas to identify gene modules defining homeostatic vs. disease-associated microglia and to evaluate their respective links with large-scale genetics, proteomics, and metabolomics data. Two canonical gene markers of homeostatic vs disease-associated microglia, namely *TMEM119* [407] for homeostatic microglia and *TREM2* [408] for disease-associated microglia, were entered as single query genes. We further analyzed the corresponding co-expression networks identified in brain tissue using the AD Atlas with regard to i) identity and number of genes that are associated with AD susceptibility and/or AD-related traits in GWAS studies, ii) identity and number of genes exhibiting increased levels in AD brains at the transcript and/or protein levels, iii) identity and number of genes that are associated with the levels of metabolites. An overview of the analysis steps can be seen in Figure 5.9 A.

We identified a total of 55 genes being co-expressed with *TMEM119* but not *TREM2* in AD tissue, of which eight genes are genetically associated with AD phenotypes: *ARHGAP45, ARPC1B, ATP8B4, GPSM3, HLA-DMA, INPP5D, SPN, TMEM106A* (Supplementary Table 8.8). Out of these eight genes, two (*ARPC1B, INPP5D*) are also associated with metabolite levels at genome-wide significance (mQTL associations). These metabolites are bilirubin, biliverdin, 1-archidonoyl-GPA (20:4), unknown metabolites, and multiple andro-

genic steroids, including epiandrosterone sulfate, androsterone sulfate and dehydroisoandrosterone sulfate (DHEA-S) (Supplementary Table 8.9). Additionally, both genes show higher gene expression levels in AD (TCX, AD vs. Control). The overlapping genetic associations with AD and levels of metabolites from the androgen pathway in the *ARPC1B* locus is of particular interest. Androgens are a class of sex steroid hormones that are responsible for the development of male sex characteristics [409] and also play important roles in female reproductive function [410]. Females have a higher susceptibility to AD [244] and studies have linked age-related depletion of the androgen testosterone to an increased risk of AD in men [411, 412]. Since *ARPC1B* has been linked to the branching and motility of microglial ramifications [413], this might suggest a potential molecular relationship between androgen levels, aging and the ability of microglia to extend ramifications in the context of AD.

We also identified 64 genes co-expressed with *TREM2* but not *TMEM119* in brain tissue, of which nine genes map to AD-associated loci: *AIF1, APOC1, GAL3ST4, HLA-DPB1, HLA-DRB1, ITGAM, ITGAX, LST1, SPI1* (Supplementary Table 8.10). In comparison to the *TMEM119* module, the *TREM2* module contains a slightly larger number of genes with overlapping AD associations, including the known genetic risk factor *APOC1* which shows strong associations to a multitude of AD phenotypes (n=35). Furthermore, while both *TMEM119*- and *TREM2*-specific networks show an overall up-regulation at the transcript level (TCX), the *TREM2* subnetwork contains nine genes (*TGFBR1, SCIN, RAB32, SPP1, CAPG, PLXDC2, FERMT3, COTL1, NPC2*) where an up-regulation can also be observed at the protein level (as opposed to none for *TMEM119*), possibly indicating a higher dysregulation at the functional level. Taken together, this supports a role for *TREM2* in the engagement of microglia toward a disease-associated phenotype. From the nine genes genetically associated with AD, six (*AIF1, APOC1, GAL3ST4, HLA-DPB1, HLA-DRB1, LST1*) are also associated with the levels of at least one metabolite. These metabolites include cholesterol, N(6)-methyllysine, multiple SMs, unknown metabolites, and multiple androgenic steroids (Supplementary Table 8.11) and three (*APOC1, HLA-DPB1, HLA-DRB1*) show higher gene expression in the TCX in AD. As observed in the *TMEM119* module, one gene of the *TREM2* module, namely *GAL3ST4*, is associated with androgen metabolites. Although few data are available regarding the role of *GAL3ST4* in microglia, the gene was reported to be part of a TYROBP brain-expressed gene module crucially involved in the development of LOAD [285]. Interestingly, a metabolite-centric search for the direct subnetwork surrounding the androgen steroids associated with both *ARPC1B* and *GAL3ST4* (overlap of metabolites is seen in Figure 5.9 B) revealed an association of these metabolites with another AD-specific risk locus; *ZCWPW1/NYAP1/PILRA* (7q22.1) [278], which has been linked to myeloid enhancer activity, microglia function and neuroinflammation [414, 415].

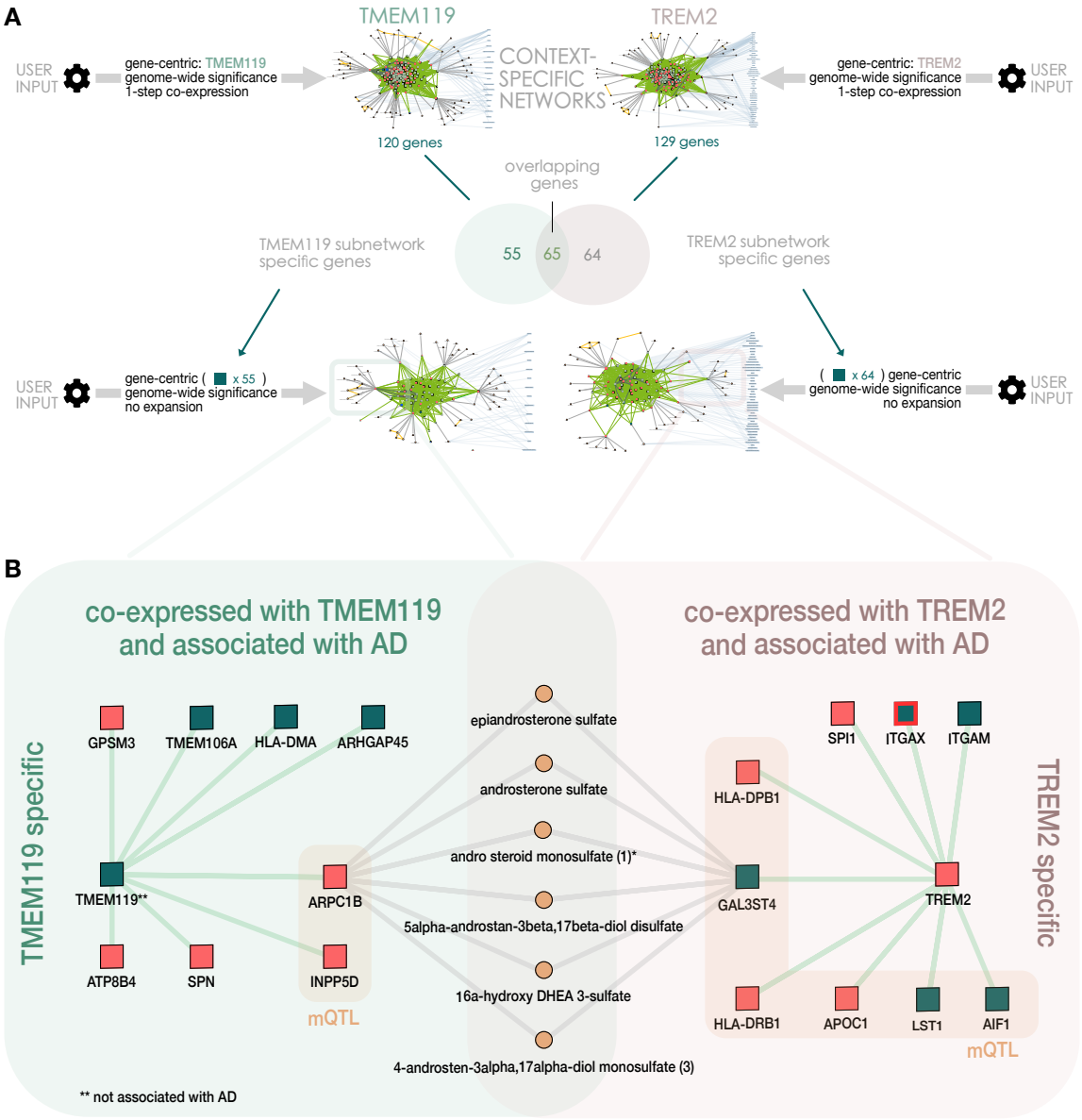


Figure 5.9: Investigating the transition between homeostatic microglia and disease-associated microglia (DAM). (**A**) Overview of the analysis. First, gene co-expression networks were built around marker genes for homeostatic (*TMEM119*) and DAM (*TREM2*). The gene modules containing genes co-expressed exclusively with one of these genes were constructed and analyzed. (**B**) Schematic representation of the marker-specific subnetworks. Genes that show co-expression in the brain with *TMEM119* but not *TREM2* and that are genetically associated with AD phenotypes are shown on the left (green background). Those that are additionally associated with metabolite levels are highlighted (orange). Genes that are significantly up-regulated at the mRNA (node filling) or protein (node border) level are highlighted in red. Genes that show co-expression in the brain with *TREM2* but not *TMEM119* and are genetically associated with AD phenotypes are shown on the right (red background). Those that are additionally associated with metabolite levels are highlighted (orange). Androgen steroids associated with both *ARPC1B* and *GAL3ST4* are depicted to highlight this overlap. It is important to note that only selected relationships are shown. For a detailed version of these networks please refer to the AD Atlas website (www.adatlas.org/?showcases).

Lastly, to see if these networks can again point to drug compounds targeting the activation of microglia, we repeated the gene set enrichment analysis described previously using the 1-step co-expression network surrounding both *TMEM119* and *TREM2*. This identified Fasudil (EnrichR, 'Drug Perturbations from GEO down', $P_{FDR} = 7.02e^{-16}$) among the top significant hits (51 unique compounds in total). This drug affects a total of 21 genes in the subnetwork and more than half are differentially expressed in AD (14 up-regulated, $P_{emp} = 0$; one down-regulated, $P_{emp} > 0.05$; Supplementary Figure 8.12). Fasudil is an inhibitor of Rho-kinase (ROCK) and is approved for the treatment of cerebral vasospasm in Japan. Post mortem data suggests that ROCK protein levels are elevated in AD brains [416] and preclinical data from *in vitro* and *in vivo* studies, including animal models of AD, indicate that Fasudil may be able to reduce the burden of tau protein [417] and promote an anti-inflammatory microglial phenotype [418, 419]. Currently, two ongoing clinical phase II trials are investigating the use of Fasudil in tauopathies (NCT04734379) and the effectiveness of an oral formulation of Fasudil in patients with dementia (NCT04793659). Interestingly, a previous study reported experimental evidence linking androgen levels through androgen receptor signaling to the levels of miRNA-135a which targets ROCK, providing a potential mechanistic model integrating the different omics entities contained in the AD Atlas-derived subnetwork [420].

Overall, our analyses point to a potential involvement of blood androgens in the transition from homeostatic to disease-associated microglia during the course of AD. Using gene set enrichment, we were able to identify Fasudil, an inhibitor of ROCK, as a promising drug repositioning candidate. We hypothesize that age-related decreases of androgen levels may result in an upregulation of ROCK via androgen-mediated pathways, leading to microglial activation and neuroinflammation [421, 422]. This exploratory analysis highlights how the AD Atlas can inform testable hypotheses, such as the potential link between androgen signaling, ROCK activity, and microglial activation in AD, that can be investigated in follow-up experiments.

Box 5.2: Definition of edge types.

In the context of this work, terms referring to different edge types of the Alzheimer's disease (AD) Atlas are defined as follows:

Co-regulation - inferred from expression quantitative trait loci (eQTL) studies to inform about shared genetic and epigenetic regulation of gene or protein expression. Edges link two or more genes mostly within one locus but can also be in trans. Interpretation examples of an edge (A)-[:COREGULATION]-(B) include:

- single nucleotide polymorphism (SNP) is an eQTL for both gene A and gene B;
- SNP is located in gene A and is eQTL for gene B;
- SNP is located in promoter/enhancer linked to gene A and is eQTL for gene B;

Genetic association - inferred from genome-wide association studies (GWAS) and metabolome-wide association studies (MWAS) to inform about genetic risk for AD and metabolite quantitative trait loci (mQTL). Edges link genes with AD-related traits or metabolic traits. Interpretation examples of an edge (A)-[:GENETIC ASSOCIATION]-(B) include:

- SNP is located in gene A and is quantitative trait loci (QTL) for (metabolic) trait B;
- SNP is an eQTL for gene A and is QTL for (metabolic) trait B;
- SNP is located in promoter/enhancer linked to gene A and QTL for (metabolic) trait B;

Metabolic association - inferred from MWAS and link metabolites with AD-related traits and indicate a significant association.

Co-expression - inferred from gene co-expression analysis and indicate a statistically significant positive or negative correlation between the levels of two transcripts.

Co-abundance - inferred from Gaussian Graphical Modelss (GGMs) using proteomics data, linking two or more genes. They indicate a significant positive or negative partial correlation between the abundance of proteins encoded by the genes.

Partial correlation - inferred from GGMs using metabolomics data and indicate a significant positive or negative partial correlation between the levels of two metabolites.

General discussion and future perspectives 6

The discussion presented in this Chapter has been partly included in the following two first-author manuscripts (*peer-reviewed journal submission in preparation*):

1. M. A. Wörheide, J. Krumsiek, S. Nataf, K. Nho, A. K. Greenwood, J. C. Wiley, T. Wu, K. Huynh, P. Weinisch, W. Römisch-Margl, N. Lehner, The AMP-AD Consortium, The Alzheimer's Disease Neuroimaging Initiative, The Alzheimer's Disease Metabolomics Consortium, J. Baumbach, P. J. Meikle, A. J. Saykin, P. Murali Doraiswamy, C. van Duijn, K. Suhre, R. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"An online molecular atlas of Alzheimer's disease"**, *submission in preparation.* Preprint available on medRxiv, doi: 10.1101/2021.09.14.21263565.
2. M. A. Wörheide, J. Krumsiek, S. Nataf, K. Nho, A. K. Greenwood, J. C. Wiley, T. Wu, K. Huynh, P. Weinisch, W. Römisch-Margl, N. Lehner, The AMP-AD Consortium, The Alzheimer's Disease Neuroimaging Initiative, The Alzheimer's Disease Metabolomics Consortium, J. Baumbach, P. J. Meikle, A. J. Saykin, P. Murali Doraiswamy, C. van Duijn, K. Suhre, R. Kaddurah-Daouk, G. Kastenmüller, and M. Arnold, **"Utilizing multi-omics context networks to explore molecular hypotheses in Alzheimer's disease"**, *submission in preparation.* Preprint available on medRxiv, doi: 10.1101/2021.09.14.21263565.

Contributions of the lead author for both manuscripts were as follows:

- implementation of the project
- data analysis and resource development (backend and frontend)
- drafting the manuscript
- design and implementation of all figures

Critical revision of the manuscripts and final approval was given by all.

6.1 Discussion

Technological advances have facilitated the generation of large-scale, high quality and high-dimensional datasets that molecularly characterize various regulatory layers of a biological system, both in health and disease. While providing exciting new opportunities to study the underpinnings of complex diseases, these developments have not come without challenges. Due to the heterogeneous and high dimensional nature of such data, interpretation

in an integrative manner, i.e., bringing together information from multiple biological domains, is not straightforward. However, for complex and multi-factorial diseases, such integrative analysis of different omics data modalities is the key to better understanding underlying molecular pathomechanisms. For example, Alzheimer's disease (AD) is a devastating neurodegenerative disorder characterized by the accumulation of neuropathological changes in the brain, which can start decades before the onset of clinical symptoms. AD remains incurable and has been linked to complex alterations in various molecular readouts, implicating biological pathways, such as membrane lipid dysregulation, impaired glucose uptake, and inflammation [19,263,277]. Therefore, integrative analyses are urgently needed to provide a more comprehensive picture and develop novel therapeutic approaches. However, there is no one-size-fits-all approach for data integration and the appropriate integration strategy often depends on the data modalities involved and whether all data was collected from the same individuals or across different cohorts. In general, two different integration strategies for the multi-omics analysis of biological datasets have been described; knowledge-based and data-driven integration.

Knowledge-based integration strategies utilize curated multi-omics knowledge bases built from large-scale experiments and scientific literature to connect different layers of biological information. For example, metabolites can be linked to enzymes/genes through pathways/reactions taken from knowledge bases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [82] or Recon3D [88]. This information can then guide multi-omics integration, for example, through set-based enrichment analysis or constraint-based metabolic models (CBMMs). However, this integration strategy is limited to characterized biological entities, and the measured entities must first be mapped to the identifier (ID) space of the knowledge base, which can lead to information loss and ambiguities. Additionally, the reported links may not be context-specific, i.e., not specific to a particular tissue or cell type. Therefore, special care must be taken when interpreting data from knowledge bases, as they may need to be filtered for information relevant to the research question (e.g., brain-specific interactions).

In contrast, data-driven strategies use statistical methods, such as correlation analysis, to infer relationships between biological entities. These methods provide context-specific insights and can include uncharacterized or only partially characterized entities in the analysis (e.g., genes with unknown functions or metabolites of unknown identity). However, this type of integration strategy can be prone to spurious associations, especially when the data is noisy, not correctly processed, or of low quality. When presented with high-quality data collected from the same set of samples, multi-block integration methods, i.e., multivariate methods that take into account all available data, have shown promising potential to help elucidate complex and multi-factorial underpinnings of diseases. However, the availability of sufficiently large cohorts characterized at multi-omics scale is currently limited. Therefore, step-wise integration methods, i.e., methods that infer relationships

between biological entities in multiple, distinct analyses and then subsequently combine the results, provide an attractive alternative. A further advantage of these methods is that they, in principle, enable the integration of data across cohorts/studies. In addition, large-scale composite network integration strategies have recently emerged. In this type of integration strategy, heterogeneous networks are constructed by inferring relationships between biological entities using knowledge-based or data-driven methods, and data are integrated by overlapping entities across omics layers (e.g. using quantitative trait loci (QTL)).

To enable an integrative analysis of AD, the aim of this thesis was to develop a network-based resource – the AD Atlas – that consolidates experimental data from large cohort studies, building a global view on molecular changes observed in AD. Through a browser-based interface, the AD Atlas provides researchers of varying backgrounds and interests an easy-to-use platform for AD knowledge discovery.

The framework underlying the AD Atlas combines multiple of the introduced multi-omics integration approaches, integrating data in a largely hypothesis-free and data-driven manner through step-wise integration combined with a composite network approach. First, we derive a molecular network by using large-scale population-based data and genomic annotation databases to provide a generalized framework to study multi-omics relationships globally. Then, we use association data from large-scale studies on diverse aspects of AD, including hundreds to hundreds of thousands of individuals, to transform this framework into an integrated multi-omics knowledge base for markers of AD. These data include recent large genetic association (meta-)analyses of AD and AD biomarkers from the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) and other large efforts, as well as differences in transcriptomics, proteomics, and metabolomics markers observed in AD or in relation to AD (endo)phenotypes. The latter are predominantly based on data generated on thousands of human blood and brain tissue samples from multiple brain regions using state-of-the-art technologies and analyzed using standardized processing pipelines through the Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) program and partnering initiatives. By integrating multiple independent datasets of the same analysis type (e.g. genome-wide association studies (GWAS) on the same traits conducted in different cohorts), we build additional data confidence by independent replication and make this data available to the user.

Building upon and extending previous efforts that have provided multi-omics integration solutions for multi-disease drug repositioning [204,312], genome-guided computational analysis of AD [188], and multi-omics annotation of individual targets [310], the AD Atlas provides users with an easy-to-use and fully network-based research platform. Through its network-focused user interface, available at www.adatlas.org, the AD Atlas enables users to conduct flexible and context-specific analyses tailored to the research question at hand without the necessity of local bioinformatics capacities. We provide multiple entry

points, allowing researchers to dynamically generate multi-omics subnetworks surrounding any gene(s), metabolite(s), or trait(s) of interest. Thereby, the analysis is not limited to the immediate multi-omics context of an individual entity. Multiple entities can be analyzed together, potentially revealing non-trivial connections, and these networks can be expanded to explore their functional neighborhood (for genes defined by co-expression, protein co-abundance, or co-regulation networks, and for metabolites by partial correlation networks). By providing additional filtering options, such as restricting co-expression links to specific brain regions, filtering for a particular sample type, or applying different significance cutoffs users can create and explore highly context-specific networks that integrate results from various sources. Besides dynamic generation and interactive exploration of networks, we additionally interlink entities to external databases and provide downstream analysis tools, including the overlay of experimental data on differentially expressed genes and differentially abundant proteins, as well as gene set and pathway enrichment analysis.

The selection of an appropriate significance cutoff is extremely important but complex. The chosen p-value cutoff greatly influences the generation of multi-omics networks as it determines which associations are represented as an edge. For the web interface, we decided to apply minimal edge filtering criteria, i.e., using genome-wide, gene-wise, or study-specific significance thresholds. While the underlying database contains potentially relevant evidence at an even finer granularity (up to a raw p-value ≤ 0.05), this data is currently not accessible through the web interface. This mixture of significance thresholds admittedly is an imperfect solution. However, we believe it represents an acceptable trade-off between inflation of false positive and false negative findings, as, in contrast to the established threshold for genome-wide significance, a similar omics-wide threshold is not available for, e.g., proteomics or metabolomics. This is due to the varying coverage of currently available platforms, with each providing measurements on only a fraction of the respective -ome. The global correlation structure of those omics thus remains unknown, and with it, the adequate number of total independent tests. Hence, applying a strict, global significance threshold based on, e.g., all known entities would lead to an inflated false negative rate. In contrast, including all nominally significant associations would minimize the false negative rate but inflate the false positive rate. Therefore, we choose to apply study-specific significance thresholds that limit false positives to 5% per study and false negatives to the rate defined by each study's power. We also do not compare or meta-analyze additional information, such as sample size, correction for confounders, or number and consistency of effects reported in the included studies. This aims at avoiding overfitting due to, for example, overlapping samples between reports or differential weighting of over-/under-studied entities, as we include associations independent of whether they have been reported in one or multiple studies. The AD Atlas thus provides a comprehensive catalog of reported associations in a network context without further judging the originally reported findings.

To build confidence in the relevance of the derived multi-omics networks and exemplify the utility of the resource to generate additional molecular hypotheses, we apply different knowledge-based assessment strategies. First, we provide evidence that the multi-omics network underlying the AD Atlas globally captures AD-relevant biology by projecting the network structure into a lower-dimensional space via node embedding and integrating this data view with established AD-related gene functions and differential expression data. We find that many clusters in this embedding space are significantly enriched for AD-related functions, including immune response, and observe distinct patterns of differential gene expression. These results suggest that functionally related genes are located in proximal regions of the network and that this association-based network is able to capture biologically relevant information. These findings are in line with results from a recent study that used the heterogeneous network-based dataset HENA [188]. This dataset integrates AD- and brain-specific data from public databases, such as Ensembl [332] and IntAct [423], and uses statistical analysis to infer interconnections from quantitative AD omics datasets. The authors subjected the resulting heterogeneous network structure to deep learning methods and showed that network topological features (e.g., graph embeddings and AD-relevant information in network neighborhoods) can provide additional information to biological features (e.g. differentially expressed genes and brain-region specific expression levels) for node classification tasks such as the prediction of AD-related genes [188]. Their analysis suggests that the topological structure of such a biological network can capture complex relationships and disease-relevant patterns that can be exploited by machine-learning approaches. In contrast to the AD Atlas, however, this dataset does not include metabolomics data and was designed for computational analyses (e.g. using machine learning methods) without offering additional tools for AD researchers lacking sufficient (bio)informatics expertise.

While such global analyses are essential to assess how genome-scale omics association data reflects known functional relationships, they are computationally expensive, and interpretability at a more fine-granular level can be challenging. This is due to both data complexity and method-dependent variation. For example, the class of embedding algorithms used in this work is based on random walks, which are not a deterministic measure of graph proximity [366]. Therefore, the embedding vectors they produce will be slightly different for each run and fine-granular details (e.g., placing of nodes in the embedding space) may change. Similarly, the dimensionality reduction method Uniform Manifold Approximation and Projection (UMAP) [351] is sensitive to parameter settings (e.g., metric: euclidean vs. cosine) and cannot fully preserve the global structure of the data [351]. Therefore, the resulting 2 dimensional (D) visual representation (and clustering) may change depending on the parameters used. Consequently, we performed our clustering analysis on the complete set of dimensions returned by GeneWalk [350] using the 2D data representation merely as a tool to present the results visually. Nevertheless, such embedding approaches offer powerful new opportunities for graph visualization and downstream analysis of complex

networks, including clustering and link prediction, which can be explored further in the future.

Next, we present multiple showcases that demonstrate the flexibility of the AD Atlas to embark on a wide variety of research questions, from drug repositioning efforts to exploratory analysis and contextualization of AD findings, and again compare these results with findings reported in the literature. For example, by investigating AD-associated genes, as well as those targeted by drug repositioning candidates proposed in the literature, we highlight the potential of the AD Atlas to aid the refinement of drug repositioning. Using two established AD risk genes, *APOE* and *CLU*, as entry points, we were able to reproduce previous research by identifying multiple repositioning candidates that are currently being tested in clinical trials, including Candesartan and Levetiracetam. In addition, we identified various other candidates that may be equally promising. We show that even without adding pathway information, extensive data on drug effects, or detailed information on effect directions, the context-specific network generated by the AD Atlas can capture disease-relevant information and guide drug prioritization. Information on differentially expressed entities (differentially expressed gene (DEG)/differentially abundant protein (DEP)) can additionally be overlaid to explore the extent of dysregulation and prioritize drug repositioning candidates that are known to induce transcriptional effects opposing the observed direction in disease [307]. In another example, we explored the proposed benefits of statins for AD therapy. In contrast to the previous analysis, where we used AD-associated genes to identify repositioning candidates, we here use the AD Atlas to evaluate a previously proposed class of drugs. By using genes that are known to be targeted by statins and evaluating them in a multi-omics disease context, we identified a link to neuroinflammation for a subset of these compounds targeting *ITGAL*. In addition to this hypothesis-driven approach to drug repositioning, we showed that the AD Atlas also allows the generation of data-driven disease networks that allow a more global and unbiased view of the disease, again identifying repositioning candidates that have been or are currently being tested in clinical trials.

Beyond applications related to drug discovery, the AD Atlas provides a valuable platform to contextualize prior analysis results and hypotheses, by enabling researchers to easily validate and complement their findings with information from multiple heterogeneous resources and publications within the scope of one resource. For example, we extend the findings of a previous analysis that implicated three sphingomyelins in AD by adding additional layers of multi-omics data to gain further insights into the potential involvement of these metabolites in AD. We find evidence that implicates the sphingosine-1-phosphate (S1P) pathway, replicating results of a recent multi-omics study. In another analysis, we relate transcriptome-derived pseudotime estimates to additional omics layers by using metabolites as proxies. This provided an extended view of the multi-omics context of metabolic correlates of pseudotime and offered further insights into the pathways underlying the

progression of AD, pointing to neurotransmission, bioenergetics, and transport-related processes.

Lastly, we showcase how the AD Atlas can be used to inform novel hypotheses on established disease-associated mechanisms. We generate subnetworks surrounding marker genes for homeostatic microglia and disease-associated microglia to investigate the underlying molecular mechanisms of this transition and identify possible drug-repositioning candidates. By focusing on genes that are genetically associated with AD phenotypes and metabolite levels, and additionally are specific to the respective marker-gene networks, we found a potential link to blood androgens, suggesting a role for these metabolites in the transition of microglia to a disease-associated state in AD. Additional gene set enrichment analysis of the molecular network surrounding both marker genes was able to identify a promising drug-repositioning candidate and provided further insights into the potentially underlying molecular mechanisms. The identified drug, Fasudil, is an inhibitor of Rho-kinase (ROCK), a protein that has been detected at elevated levels in the post-mortem brains of AD patients. Experimental evidence suggests that decreased androgen levels, as observed in aging, may result in an up-regulation of ROCK via androgen-mediated pathways and lead to neuroinflammation. Therefore, suggesting a potential interplay between androgen signaling and Rho kinase activity in AD.

While this work is focused on AD, the framework that underlies the AD Atlas can easily be adapted or extended to other disorders. The technical framework was built to allow the continuous extension of the resource both in regards to additional datasets of known formats as well as the addition of novel data modalities. Many incorporated datasets are not disease-specific, such as genetic associations with metabolic traits that have been inferred from population-based cohorts and represent general molecular pathways. Furthermore, disease-specific datasets, such as brain co-expression networks, can be replaced by expression networks from tissue(s) relevant to the respective disease. For example, when building an atlas of chronic kidney disease, integrated datasets could include kidney co-expression and protein co-abundance networks, urine metabolomics studies, and GWAS with biomarkers of chronic kidney disease. Lastly, by integrating association results across potentially related diseases, such as AD and type 2 diabetes - which is considered a risk factor for AD, this framework can facilitate the exploration of shared disease mechanisms.

6.2 Limitations and future perspectives

The AD Atlas in its current form provides extended access to large-scale multi-omics data generated by interdisciplinary AD initiatives, embedding single biological entities into their molecular multi-level context. As such, in its current version, the resource has several (inherent) limitations but also provides an abundance of opportunities, which I will discuss in the following.

First, based on the currently limited availability of research data obtained in diverse populations/ethnicities, the data used to generate the AD Atlas are also not representative of these populations, which reduces the generalizability and transferability of the integrated molecular observations. However, the AMP-AD 2.0 program has the particular mission to expand research efforts to Latino and African American communities, and we will put a particular emphasis on the integration of resulting findings from these diverse populations in the AD Atlas resource. By adding data from different populations, we can compare and contrast molecular effects observed in AD. This will allow us to differentiate between population-specific and common effects, with the latter being prime candidates for targeting (one drug to cure them all).

Second, while the AD Atlas framework integrates metabolomics data at the measured compound level, i.e., uses platform-specific identifiers to link results where possible, not all studies report this level of information. When only biochemical names are provided, this may result in information loss and ambiguity, as the level of resolution can vary greatly between different platforms. Therefore, consolidating results across platforms based on their given biochemical label is not optimal, especially in extreme cases where the reported label corresponds to unrelated signals on different platforms. To circumvent this problem, we only consolidated cross-platform metabolites based on manual mappings and are currently working towards classifying measured metabolites in a manually curated ontology that can accommodate the differing levels of technological resolution and reporting. This would lead to a more precise representation of metabolomics results, allowing users to explore results at different levels of granularity and pinpoint disease-associated species. This is of great importance as, for example, lipid side chain composition and configuration can be important determinants of metabolite function.

Third, the application of the AD Atlas for drug repositioning presented in this work shows promising potential. To realize the full potential of the AD Atlas, we plan to develop more sophisticated drug repositioning algorithms and incorporate brain-specific protein-protein interaction (PPI) networks and cell-type specific drug perturbation signatures. To date, many transcriptional signature databases, such as the Broad Institute Connectivity Map (CMap) [424] or LINCS L1000 [425], are heavily focused on cancer cell lines, with limited neuronal or brain-relevant information. However, as more brain- and cell-type specific information becomes available (e.g., through scRNA-seq studies [426]), this data can be integrated into the AD Atlas, providing exciting possibilities.

Fourth, omics data from tissues other than blood and brain are currently underrepresented in the AD Atlas. To address this, we are actively working on integrating data on a broader set of tissues to comprehensively account for risk factors for AD. For instance, assessing the effects of environmental exposures, lifestyle, and gut microbiome composition on AD necessitates the inclusion of additional peripheral tissues and biofluids such as urine (e.g., for measuring toxins or kidney function) and feces (e.g., for changes in gut microbiome

metabolism). Nevertheless, because tissue specificity and the focus on molecular changes in the brain are important aspects to consider when studying AD, we provide tools to filter results to tissue-specific relationships. However, the expansion to more comprehensive data from the periphery and the combined analysis of omics data across tissues holds the potential to uncover so far unknown interactions between environmental factors, other organ systems, and AD pathogenesis in the brain.

Lastly, we have mainly focused on developing tools for exploring hypothesis-guided local subnetworks. However, the network underlying the AD Atlas has great potential to identify disease modules globally in a more hypothesis-free manner. As we show, analyzing the whole network via node embeddings provided promising first results, and we are currently working towards extending these efforts to make more powerful analysis tools available through the AD Atlas web interface.

7 Conclusion

In the face of technological advances and falling costs for high-throughput profiling technologies, the integrative analysis of multi-modal biological data has become an attractive strategy to complement traditional omics data analysis. Especially for complex and untreatable diseases linked to changes across omics layers, such as Alzheimer's disease (AD), integrative studies can provide a more comprehensive view of the disease, deepen our understanding of underlying molecular mechanisms and help uncover novel therapeutic targets. However, the growing complexity (high dimensionality, increasing data heterogeneity) and abundance of these datasets provide novel challenges both computationally and conceptually (interpretability of complex and context-specific interactions). Large-scale integrative analysis requires (bio-)informatics expertise and often involves multiple, time-consuming analysis steps which include identifying the appropriate integration strategy (no 'one-size-fits-all' solution) and acquiring and preprocessing the data (e.g., data access restrictions, identifier mapping). Therefore, user-friendly and flexible frameworks that facilitate the exploration of multi-omics data for researchers of varying backgrounds are needed.

To address this need, I first provide an introduction to large-scale multi-omics data integration strategies, focusing on methods that enable the integration of metabolomics data with other omics datasets (**Scientific aim 1**). This detailed overview provides a general guide for researchers interested in integrative analysis to identify appropriate integration strategies for their research question and data sets at hand (**Chapter 2**). Next, I introduce the complex neurodegenerative disorder AD and present the current understanding of disease pathophysiology that has been linked to molecular changes across omics layers (**Chapter 3**). To enable the integrative analysis of molecular data generated and made available by international and interdisciplinary AD initiatives (**Scientific aim 2**), I then present a global framework that integrates different omics datasets across studies by building an extensive heterogeneous network of molecular interconnections (**Chapter 4**). By applying this framework, I created the AD Atlas resource (**Chapter 5**) – a network-based, online multi-omics platform that provides access to a wide variety of population-based and AD-specific data. To provide researchers access to this resource, regardless of their computational expertise (**Scientific aim 3**), I built a network-based user interface that enables

users to view biological entities of interest, i.e. metabolites, genes or AD (endo)phenotypes, in an AD-relevant, multi-omics context. In addition, downstream analysis tools provided by the AD Atlas enable complex multi-omics analyses that can be performed with just a few clicks. In multiple showcases, I provide evidence that the underlying network of inter- and intra-omics links provides biologically relevant information and can capture (known) AD-relevant facts. Moreover, these examples demonstrate the usability and relevance of this resource for AD research (**Scientific aim 4**). Lastly, I discuss the potential and limitations of this new resource and provide future directions for the presented work, including the application of this integration framework to other complex diseases (**Chapter 6**).

While preprocessing and analyzing complex biological data will most likely remain tedious and time-consuming, integrating and sharing such data at a global scale can provide countless novel opportunities for downstream analysis. Especially by modeling the data and relationships as networks, several established graph algorithms can be applied. Furthermore, by not only making such data collections available in supplementary tables but also providing access and interpretation through user interfaces that facilitate complex analysis, the utility of large-scale studies and analyses can be further extended. Therefore, I firmly believe that the effort required to develop and implement such platforms is offset by the value the resulting resources provide to the research community.

In conclusion, this thesis addresses multiple bottlenecks in the translation of multi-omics data into disease-relevant insights by (i) providing a comprehensive guide to current integration strategies, (ii) developing a global framework for the integrative analysis of complex diseases, and (iii) applying this framework to build a network-based online research platform for AD. The thereby developed AD Atlas is a unique multi-omics resource that provides an interpretable, global view of highly complex biological relationships in the context of AD. Through a user-friendly interface, the resource enables users to perform tailored analyses and formulate molecular hypotheses that can be validated in follow-up analyses or experiments. While the application examples presented provide compelling showcases of what this resource is capable of, I see the true potential of this work in the variety of research questions that can be explored and the flexible application opportunities this framework offers for future work and other complex diseases.

Publication record

Peer-reviewed journal articles

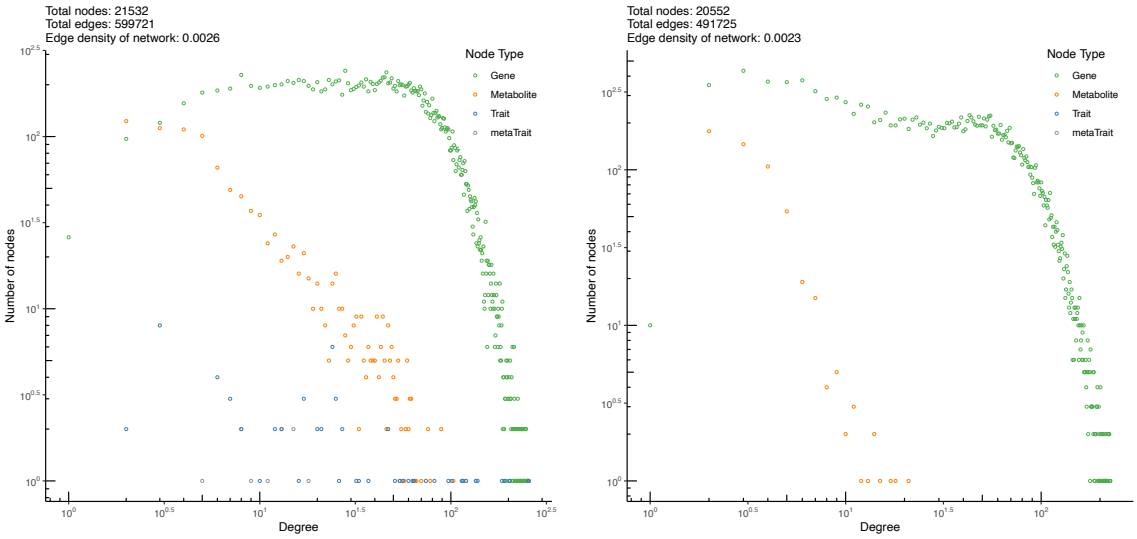
1. P. Surendran, I. D. Stewart, V. P. W. Au Yeung, M. Pietzner, J. Raffler, M. A. Wörheide, C. Li, R. F. Smith, L. B. L. Wittemans, L. Bomba, C. Menni, J. Zierer, N. Rossi, P. A. Sheridan, N. A. Watkins, M. Mangino, P. G. Hysi, E. Di Angelantonio, M. Falchi, T. D. Spector, N. Soranzo, G. A. Michelotti, W. Arlt, L. A. Lotta, S. Denaxas, H. Hemingway, E. R. Gamazon, J. M. M. Howson, A. M. Wood, J. Danesh, N. J. Wareham, G. Kastenmüller, E. B. Fauman, K. Suhre, A. S. Butterworth and C. Langenberg, **"Rare and common genetic determinants of metabolic individuality and their effects on human health"**, *Nat. Med.*, vol. 28, pp. 2321-2332, Nov. 2022.
2. R. Batra, M. Arnold, M. A. Wörheide, M. Allen, X. Wang, C. Blach, A. I. Levey, N. T. Seyfried, N. Ertekin-Taner, D. A. Bennett, G. Kastenmüller, R. F. Kaddurah-Daouk, J. Krumsiek, and Alzheimer's Disease Metabolomics Consortium (ADMC), **"The landscape of metabolic brain alterations in Alzheimer's disease"**, *Alzheimers. Dement.*, pp. 1-19, doi: 10.1002/alz.12714, July 2022.
3. M. Pietzner, E. Wheeler, J. Carrasco-Zanini, A. Cortes, M. Koprulu, M. A. Wörheide, E. Oerton, J. Cook, I. D. Stewart, N. D. Kerrison, J. Luan, J. Raffler, M. Arnold, W. Arlt, S. O'Rahilly, G. Kastenmüller, E. R. Gamazon, A. D. Hingorani, R. A. Scott, N. J. Wareham, and C. Langenberg, **"Mapping the proteo-genomic convergence of human diseases"**, *Science*, vol. 374, p. eabj1541, Nov. 2021.
4. M. A. Wörheide, J. Krumsiek, G. Kastenmüller, and M. Arnold, **"Multi-omics integration in biomedical research – a metabolomics-centric review"**, *Anal. Chim. Acta*, vol. 1141, pp. 144–162, Jan. 2021.

Peer-reviewed conference proceedings

1. M. A. Wörheide, G. Kastenmüller, R. F. Kaddurah-Daouk, M. Arnold, **"A proof of concept study towards multi-omics-based computational drug repositioning in Alzheimer's disease"**, *Alzheimers. Dement.*, Dec. 2021.
2. M. A. Wörheide, J. Krumsiek, K. Nho, K. Huynh, P. J. Meikle, A. J. Saykin, R. F. Kaddurah-Daouk, G. Kastenmüller, M. Arnold, **"A network-based, multi-omics atlas for target identification and prioritization in Alzheimer's disease"**, *Alzheimers. Dement.*, Dec. 2020.

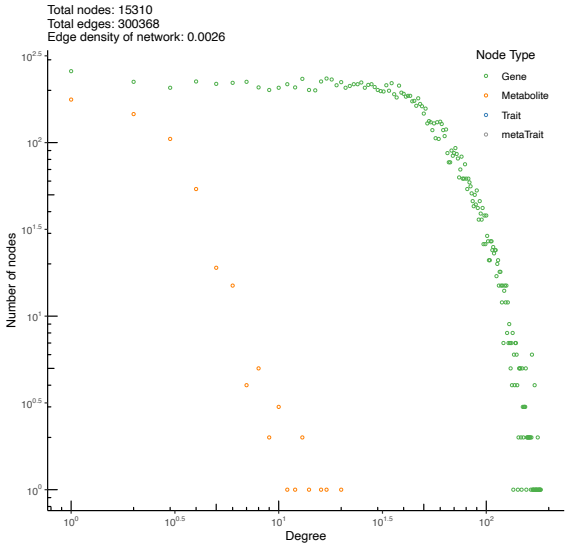
Appendix 8

Figure 8.1: Example cypher query. Users can dynamically create context-specific networks by using the 'Network settings' sidebar of the AD Atlas web interface. The user-specified options are then translated internally into a cypher query that is used to communicate with the underlying Neo4j database. The returned data matrices are then processed, which includes further edge filtering, and used by the AD Atlas frontend to visualize the interactive multi-omics network using the R package *visNetwork*.



(A) all

(B) brain-specific



(C) brain-specific and excl. co-regulation

Figure 8.2: Degree distribution of the AD Atlas network. (**A**) Simple network representation (without self-loops and multiple edges between nodes) of the AD Atlas including all edge types, sample types, and node types. (**B**) Simple network representation of the AD Atlas filtered to brain-specific edges. (**C**) Simple network representation of the AD Atlas filtered to brain-specific edges, and excluding co-regulation edges and trait nodes as used in the global network structure analysis in Chapter 5 Section 5.3.

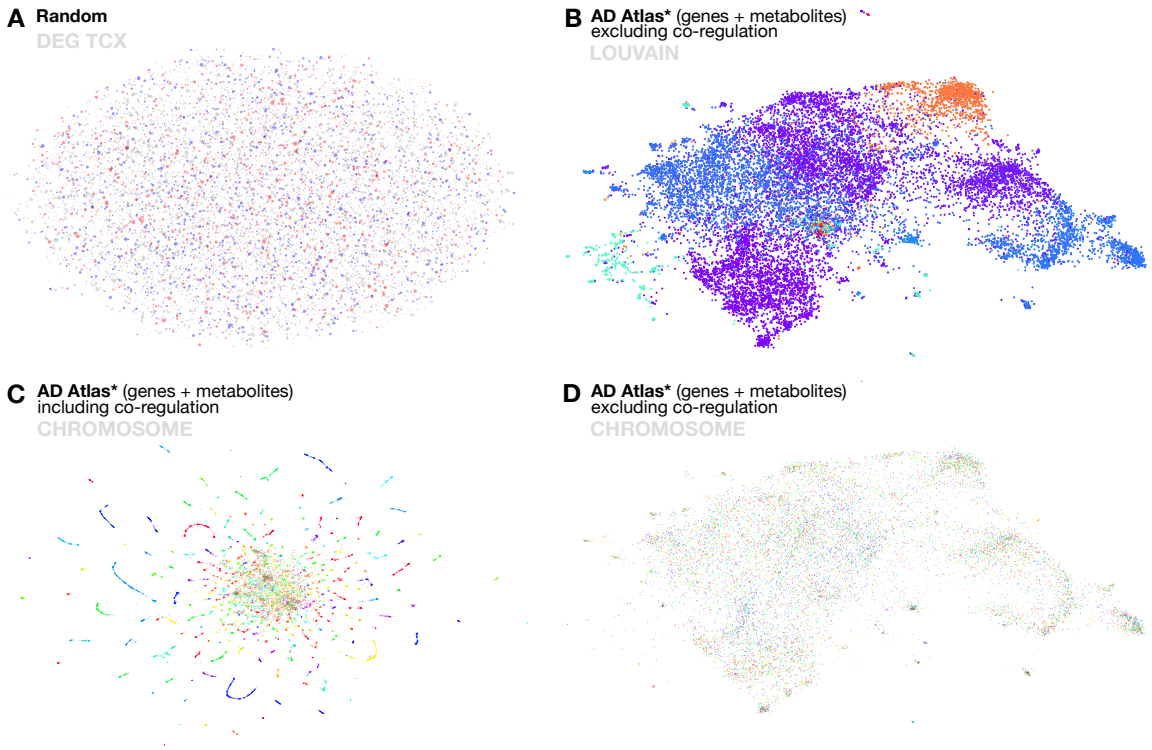


Figure 8.3: Additional analysis of the global network structure of the AD Atlas. The nodes of the AD Atlas network were embedded into an 89-dimensional vector space and projected to 2D for visualization, as described in Chapter 4 Section 4.4.1. (**A**) A randomized network as implemented by the GeneWalk [350] package, colored by differential expression in the temporal cortex (TCX). This shows, as expected, no structure or clustering. (**B**) Communities found by applying the network community detection algorithm Louvain [427] (as implemented by the Python package *NetworkX*; default settings) directly on the (simplified) network. Communities found in the network align with the overall topology of embedding visualization. (**C**) Including co-regulation edges (filtered for brain-specific expression quantitative trait locis (eQTLs)) in the network embedding introduces clusters primarily driven by genetic architecture. The genomic location of genes indicated by different colors for each chromosome. (**D**) Excluding co-regulation edges largely removes this effect with only a few chromosomal clusters remaining. Gene nodes are again colored by chromosomes. *Brain-specific.

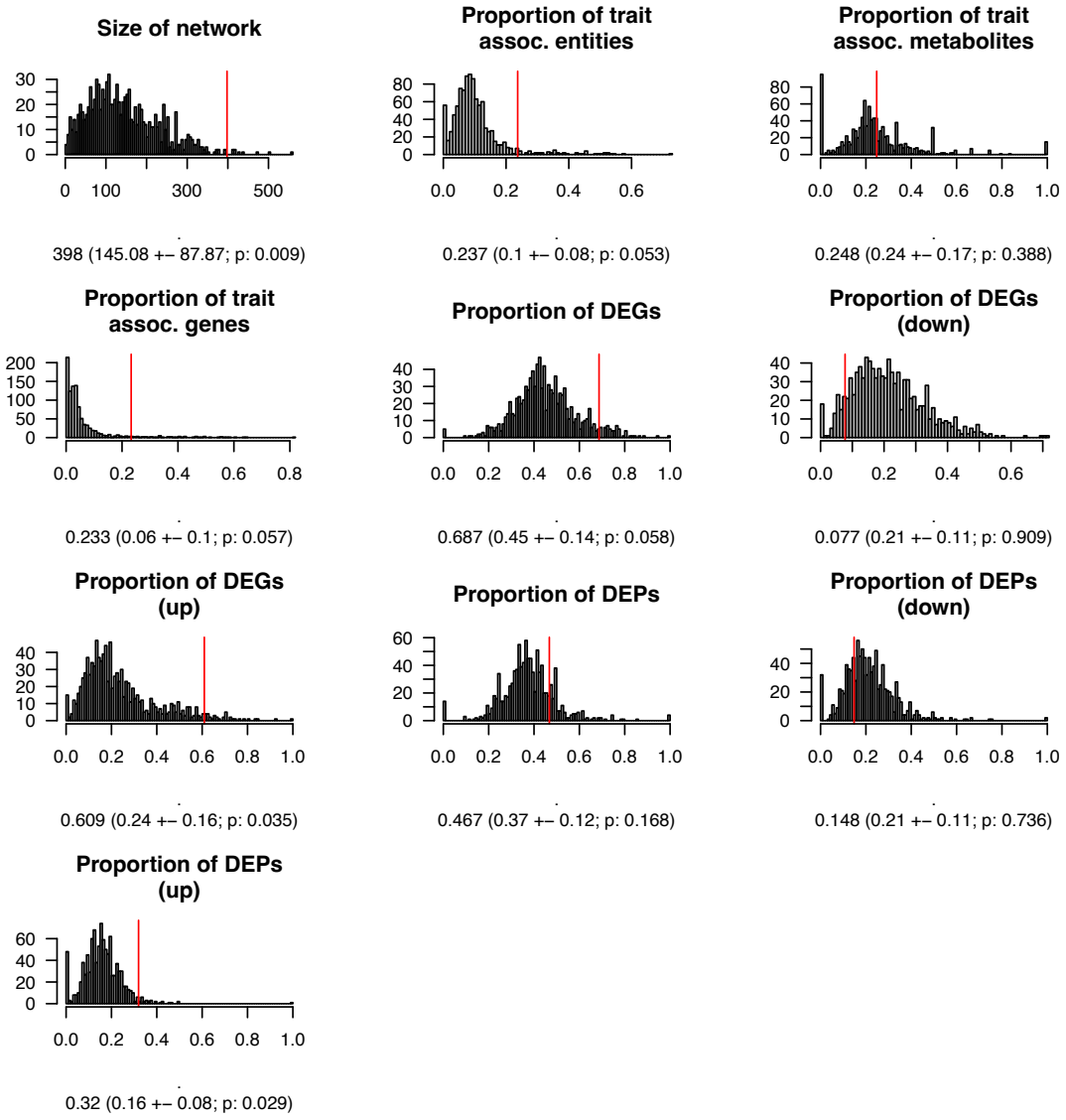
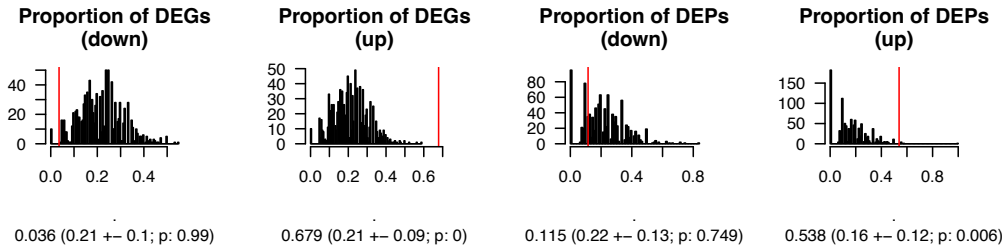
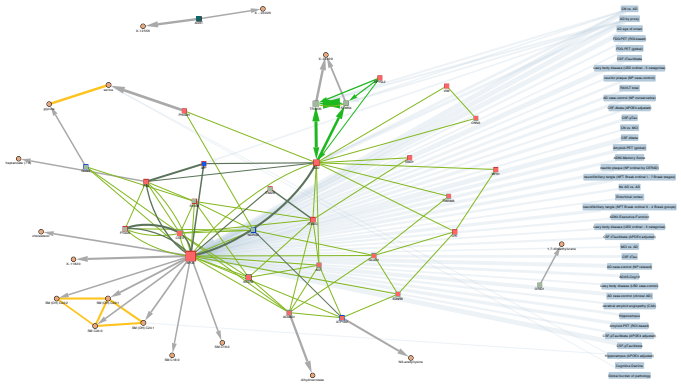
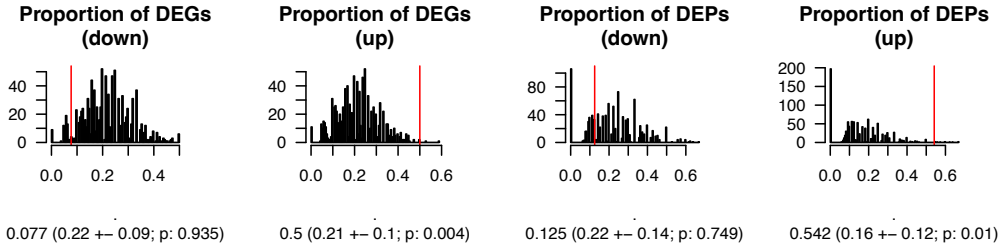


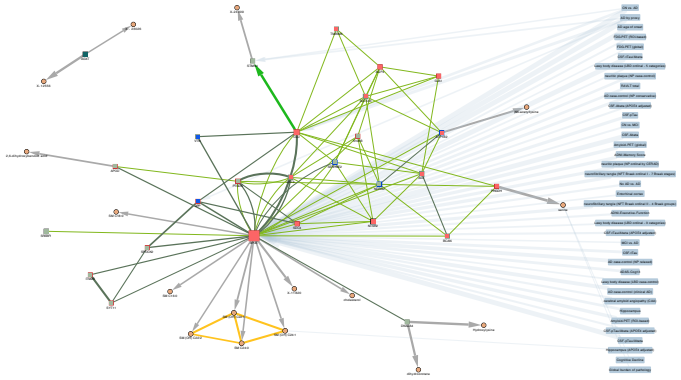
Figure 8.4: Background distributions for the $APOE/CLU$ example. Network properties of 1000 random networks built by selecting genes (n=2) and expanding to their 1-step neighborhood, filtering co-regulation edges to the brain cortex. Red line indicates the observed property for the $APOE$ and CLU molecular subnetwork (Figure 5.4 A). The observed value with mean, standard deviation, and empirical p-value is provided below each histogram.





(A) Levetiracetam





(B) Candesartan

Figure 8.5: Molecular network of genes affected by repositioning candidates. (**A**) Levetiracetam and (**B**) Candesartan. Both networks include genes from distinct genetic loci, as can be seen by using the "Node annotation – Show annotation for genes" function on the left panel of the AD Atlas. Furthermore, they are only connected by a few co-regulation links which often indicate common loci. The genes are connected at the messenger ribonucleic acid (mRNA) and protein level through co-expression (light green) and co-abundance (dark green) edges. Above each network the background distribution of the proportion of differentially expressed genes (DEGs) or differentially abundant proteins (DEPs) from 1000 random gene sets is shown (n=27 genes for Levetiracetam and n=26 genes for Candesartan). Networks were constructed using URL queries of the form: adatlas.org/?geneSymbol={gene symbols taken from EnrichR output}&gtex=Brain%20Cortex.

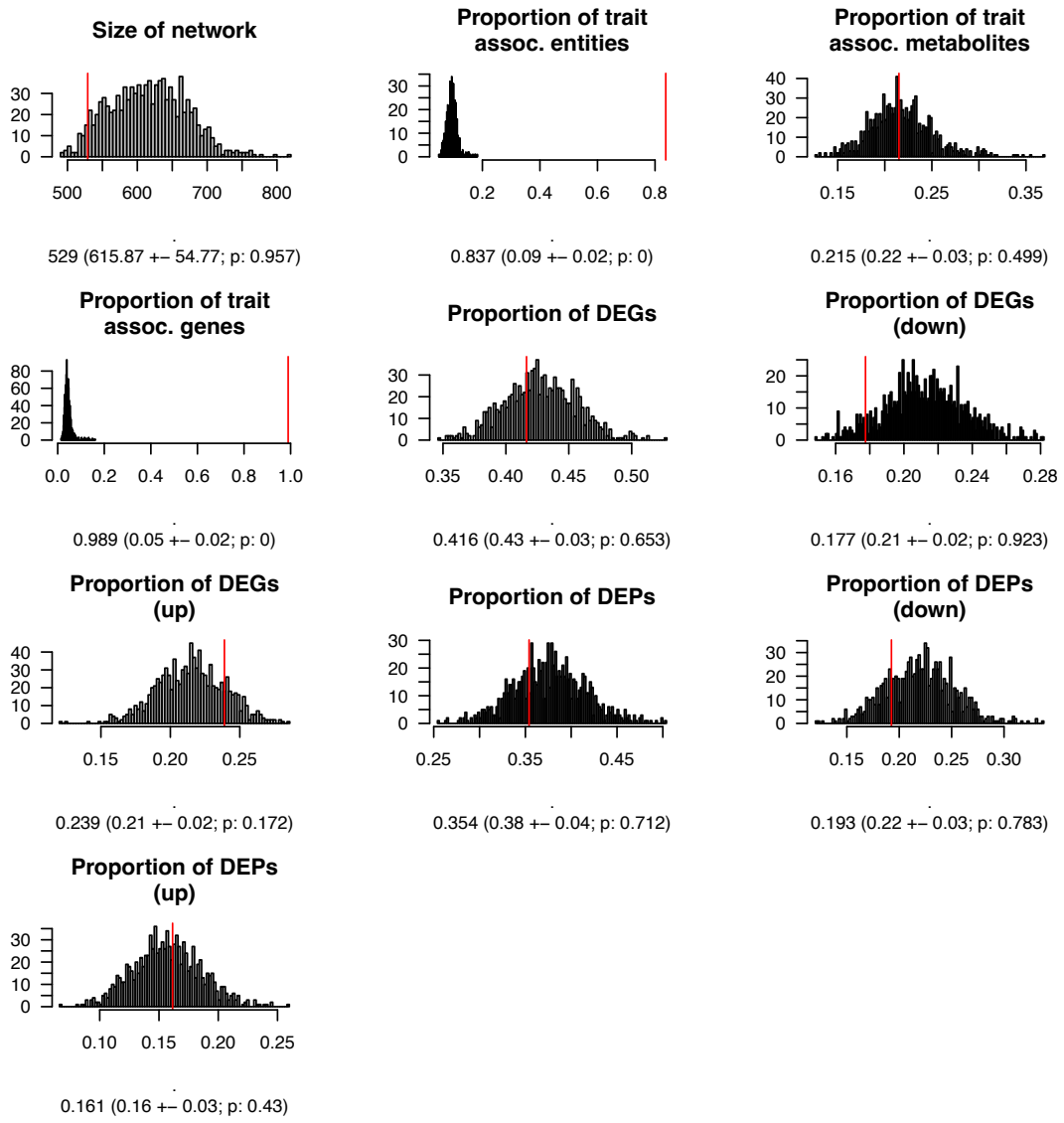
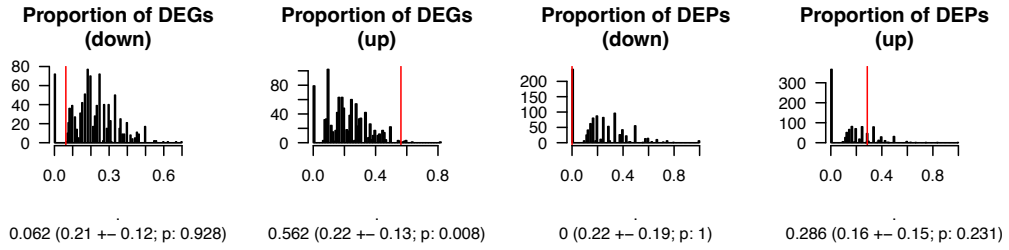
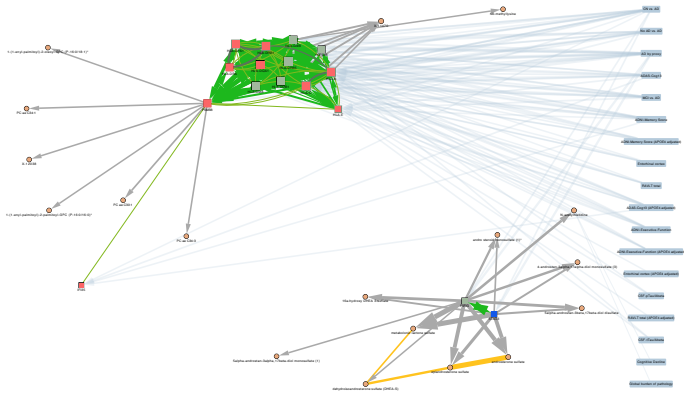
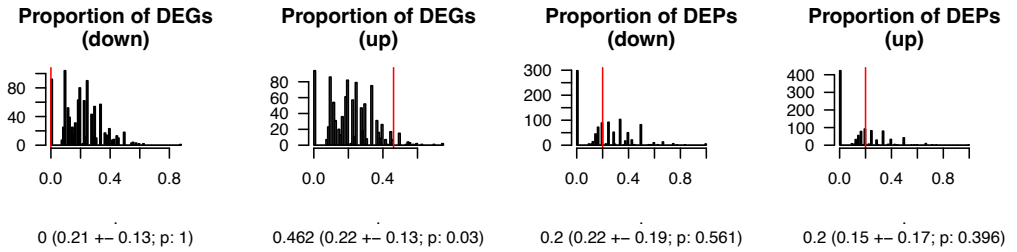


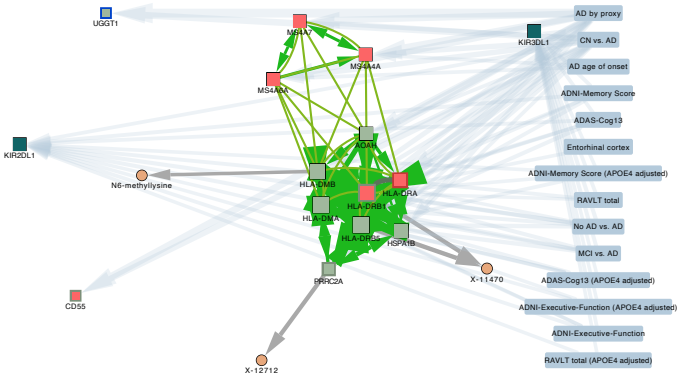
Figure 8.6: Background distributions for AD case-control example. Network properties of 1000 random networks built by selecting genes (n=376) and metabolites (n=5) and annotating these with significant associations from genome-wide association studies with metabolite traits (mGWAS), genome-wide association studies (GWAS) and metabolome-wide association studies (MWAS). Red line indicates observed property for the trait-centric case-control molecular subnetwork (seen in Figure 5.5 A). The observed value with mean, standard deviation, and empirical p-value is provided below each histogram.





(A) Citalopram





(B) Etarnecept

Figure 8.7: Molecular network of genes affected by repositioning candidates. (**A**) Citalopram and (**B**) Etanercept. Both networks primarily target one genetic locus (tightly linked clusters through co-regulation links). Above each network, the background distribution of the proportion of DEGs or DEPs from 1000 random gene sets is shown (n=16 genes for Citalopram and n=14 genes for Etanercept). Networks were constructed using URL queries of the form: adatlas.org/?geneSymbol={gene symbols taken from EnrichR output}&gtex=Brain%20Cortex.

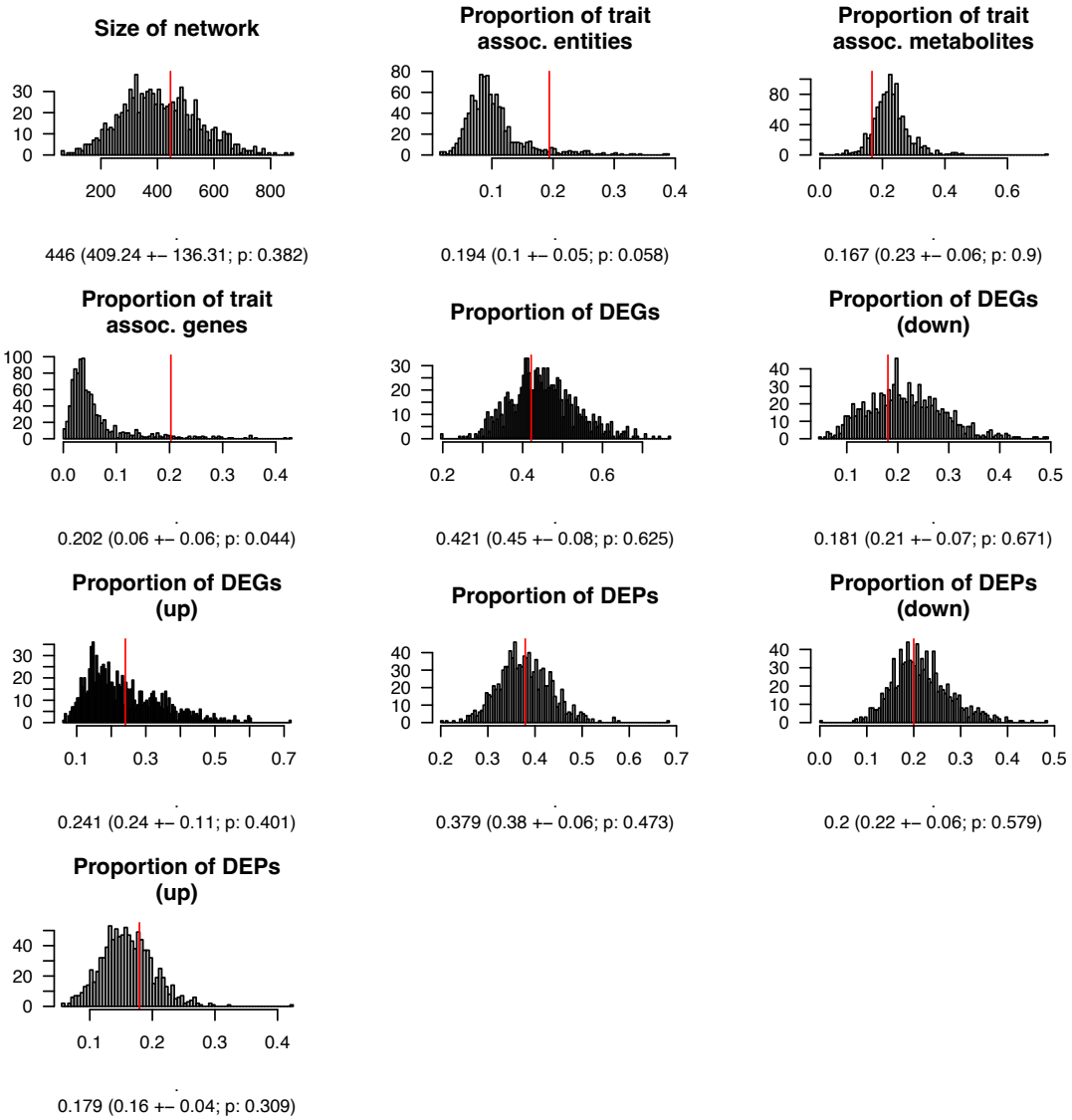


Figure 8.8: Background distribution for statins example. Network properties of 1000 random networks built by selecting genes (n=6) and expanding to their 1-step neighborhood, filtering co-regulation edges to the brain cortex. Red line indicates the observed property for the gene-centric molecular subnetwork surrounding statin targets (seen in Figure 5.6 A). The observed value with mean, standard deviation, and empirical p-value is provided below each histogram.

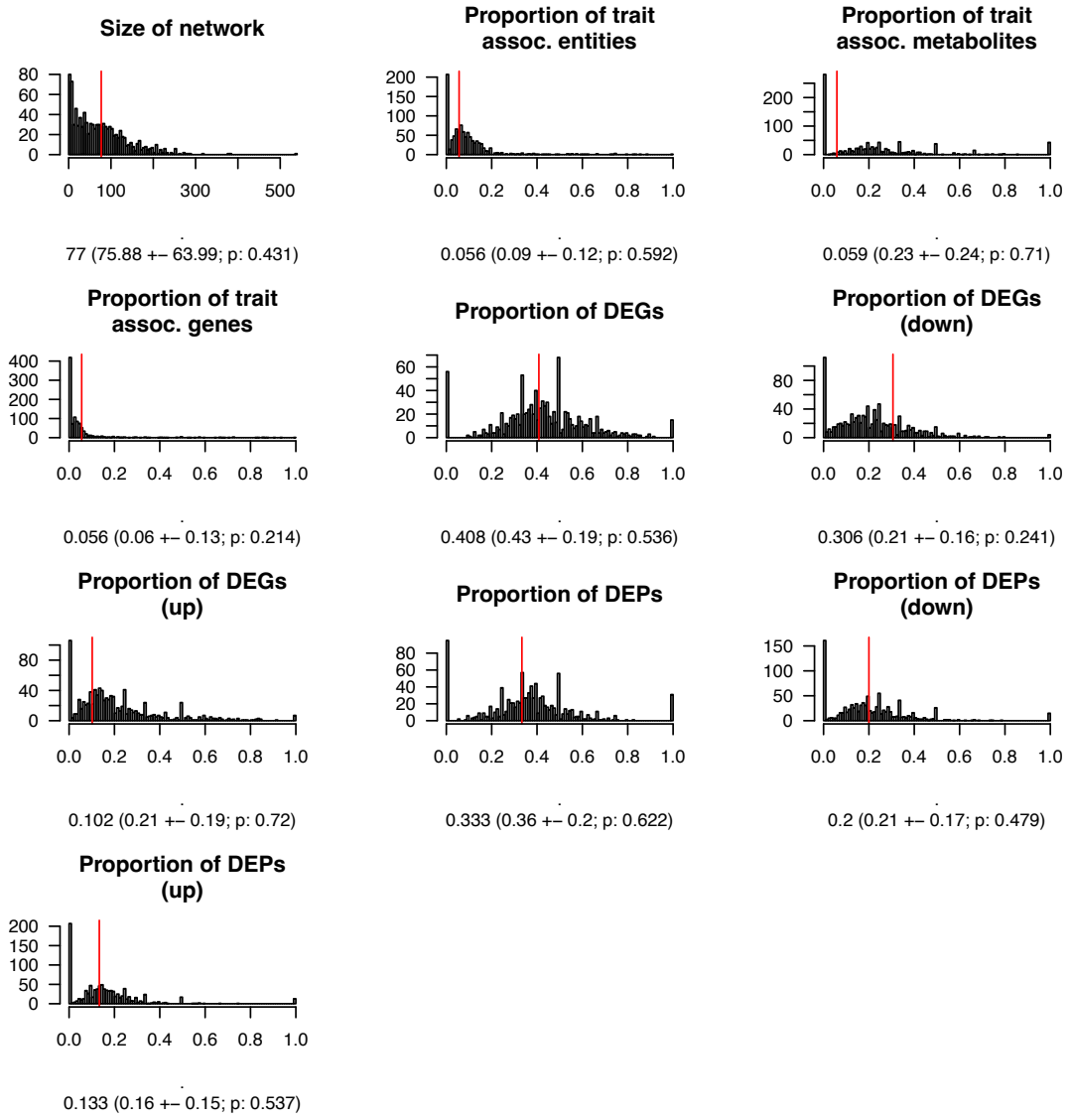


Figure 8.9: Background distribution for statins example (gene *HMGCR*). Network properties of 1000 random networks built by selecting one gene and expanding to their 1-step neighborhood, filtering co-regulation edges to the brain cortex. Red line indicates the observed property for the gene-centric molecular subnetwork surrounding the gene *HMGCR* (seen in Figure 5.6 C). The observed value with mean, standard deviation, and empirical p-value is provided below each histogram.

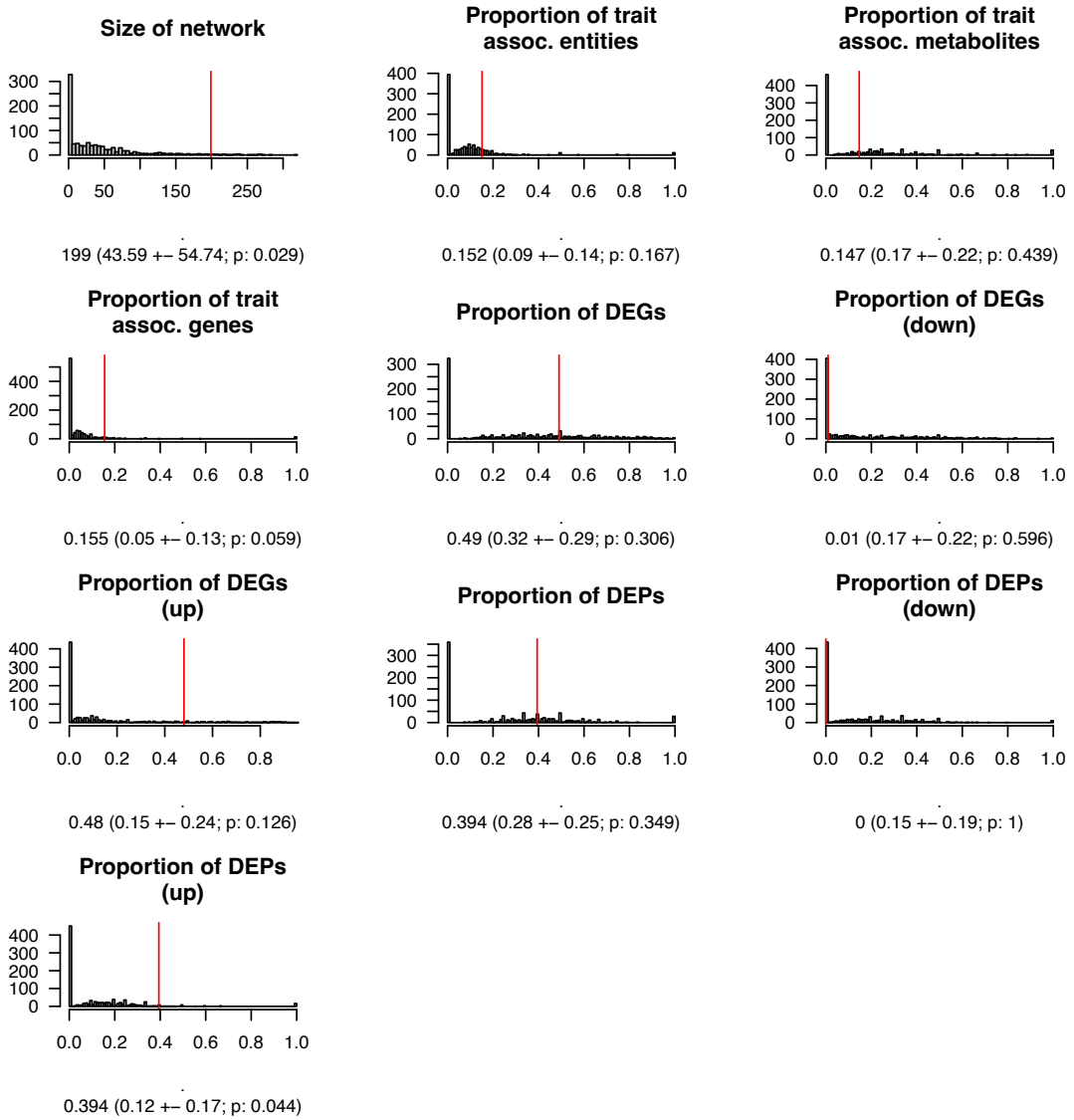


Figure 8.10: Background distribution for statins example (gene *ITGAL*). Network properties of 1000 random networks built by selecting one gene and expanding to their 1-step co-expression neighborhood. Red line indicates the observed property for the gene-centric molecular subnetwork surrounding the gene *ITGAL* (seen in Figure 5.6 B). The observed value with mean, standard deviation, and empirical p-value is provided below each histogram.

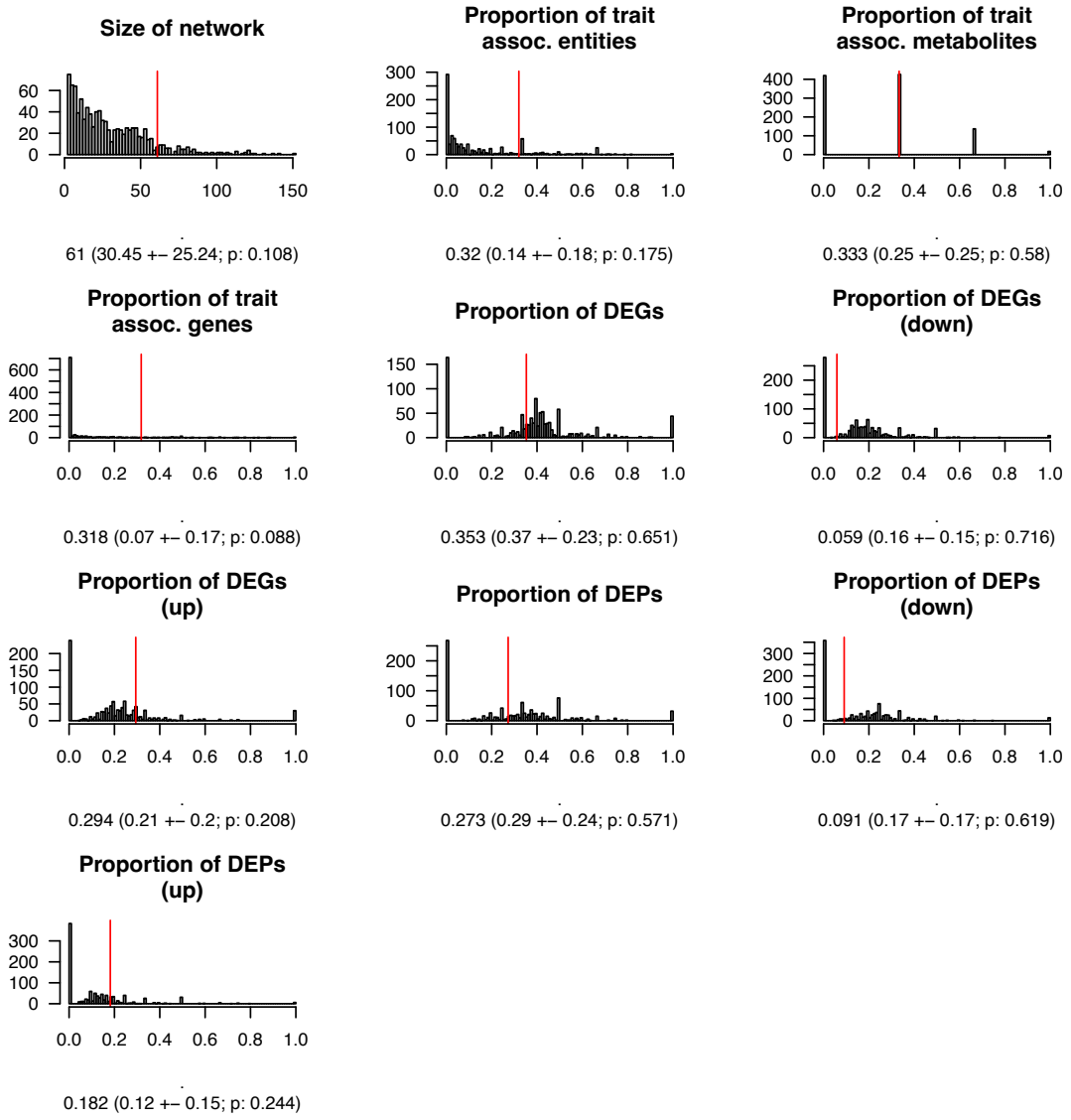


Figure 8.11: Background distribution for sphingomyelins example. Network properties of 1000 random networks built by selecting metabolites (n=3) and annotating these with significant associations from mGWAS, GWAS, and MWAS. Red line indicates the observed property for the metabolite-centric molecular subnetwork surrounding three sphingomyelins (seen in Figure 5.7 A). Observed value with mean, standard deviation, and empirical p-value are provided below each histogram.

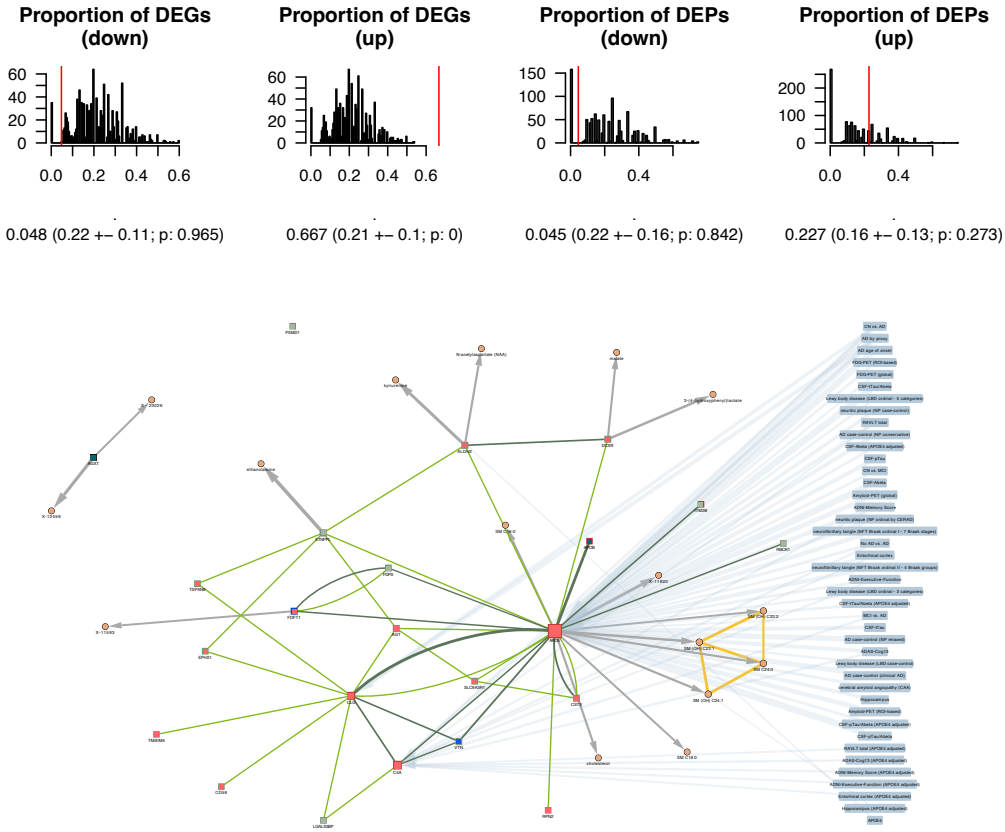


Figure 8.12: Molecular network of genes affected by repositioning candidate Fasudil. The network includes genes from distinct genetic loci, as can be seen by using the "Node annotation – Show annotation for genes" function on the left panel of the AD Atlas. Furthermore, they are not connected by co-regulation links, which often indicates common loci. The genes are connected both at the mRNA and protein level through co-expression (light green) and co-abundance (dark green) edges. Above the network, the background distribution of the proportion of DEGs or DEPs from 1000 random gene sets is shown (n=21 input genes). Network was constructed using URL queries of the form: adatlas.org/?geneSymbol={gene symbols taken from EnrichR output}&gtex=Brain%20Cortex.

Table 8.1: Data summary and statistics. Edge numbers differ from the main text table as undirected edges are modeled by two directed edges in Neo4j.

Nodes	***n***
metaTrait mt	10
trait t	67
gene (protein-coding) g	20,363
DEG in min. 1 AD phenotype	14,731
DEP in min. 1 AD phenotype	7,867
metMeta m	1,328
Edges	***n***
(t) - [:PART_OF] ->(mt)	105
(g) - [:GENETIC_ASSOCIATION] –>(t)	12,361
(g) - [:GENETIC_ASSOCIATION] – >(m)	165,719
(m) - [:METABOLIC_ASSOCIATION] –>(t)	1,018
(g) - [:COEXPRESSION] – (g)	465,032
(g) - [:COEXPRESSION {tissue:CBE}] – >(g)	75,066
(g) - [:COEXPRESSION {tissue:DLPFC}] – >(g)	106,156
(g) - [:COEXPRESSION {tissue:FP}] – >(g)	40,294
(g) - [:COEXPRESSION {tissue:IFG}] – >(g)	30,670
(g) - [:COEXPRESSION {tissue:PHG}] – >(g)	34,892
(g) - [:COEXPRESSION {tissue:STG}] – >(g)	35,656
(g) - [:COEXPRESSION {tissue:TCX}] – >(g)	64,230
(g) - [:COREGULATION] – >(g)	493,117
(g) - [:COREGULATION_{tissue}] – >(g)	4,345,276
(g) - [:COREGULATION_BRAIN_AMYGDALA] – >(g)	23,817
(g) - [:COREGULATION_BRAIN_ANTERIOR_CINGULATE_CORTEX_BA24] – >(g)	37,213
(g) - [:COREGULATION_BRAIN_CAUDATE_BASAL_GANGLIA] – >(g)	61,761
(g) - [:COREGULATION_BRAIN_CEREBELLAR_HEMISPHERE] – >(g)	78,049
(g) - [:COREGULATION_BRAIN_CEREBELLUM] – >(g)	95,397
(g) - [:COREGULATION_BRAIN_CORTEX] – >(g)	246,462
(g) - [:COREGULATION_BRAIN_FRONTAL_CORTEX_BA9] – >(g)	53,237
(g) - [:COREGULATION_BRAIN_HIPPOCAMPUS] – >(g)	38,548
(g) - [:COREGULATION_BRAIN_HYPOTHALAMUS] – >(g)	37,824
(g) - [:COREGULATION_BRAIN_NUCLEUS_ACCUMBENS_BASAL_GANGLIA] – >(g)	60,903
(g) - [:COREGULATION_BRAIN_PUTAMEN_BASAL_GANGLIA] – >(g)	50,454
(g) - [:COREGULATION_BRAIN_SPINAL_CORD_CERVICAL_C-1] – >(g)	28,965
(g) - [:COREGULATION_BRAIN_SUBSTANTIA_NIGRA] – >(g)	20,255
(g) - [:COABUNDANCE] – (g)	146,592
(m) - [:PARTIAL_CORRELATION] – (m)	2,326

Table 8.2: Collections of AD-related phenotypes summarized as 'metaTraits'. A trait can be connected to multiple meta-traits. Further trait descriptions are given in Supplementary Table 8.3.

Meta trait	Trait
AD CSF biomarker	CSF-Abeta
	CSF-Abeta (APOE4 adjusted)
	CSF-Abeta pathology (based on threshold)
	CSF-CLU
	CSF-pTau
	CSF-pTau (APOE4 adjusted)
	CSF-pTau/Abeta
	CSF-pTau/Abeta (APOE4 adjusted)
	CSF-tTau
	CSF-tTau (APOE4 adjusted)
	CSF-tTau/Abeta
	CSF-tTau/Abeta (APOE4 adjusted)
AD imaging biomarker	Amyloid-PET (ROI-based)
	Amyloid-PET (ROI-based; APOE4 adjusted)
	Amyloid-PET (global)
	Amyloid-PET (global; APOE4 adjusted)
	Cortex
	Entorhinal cortex
	Entorhinal cortex (APOE4 adjusted)
	FDG-PET (ROI-based)
	FDG-PET (ROI-based; APOE4 adjusted)
	FDG-PET (global)
	FDG-PET (global; APOE4 adjusted)
	Hippocampus
	Hippocampus (APOE4 adjusted)
	Ventricles
	White matter hyperintensities
	White matter hyperintensities (APOE4 adjusted)
AD risk/diagnosis	AD age of onset
	AD case-control (NP conservative)
	AD case-control (NP relaxed)
	AD case-control (clinical AD)
	AD by proxy
	APOE4
	Clinical diagnosis (ordinal)
	CN vs. AD**
	CN vs. MCI*
	MCI vs. AD*
	Neuropathology diagnosis
	NIA-Reagan Score
	No AD vs. AD*

Amyloid pathology	Amyloid beta (Immunohistochemistry-based) Amyloid-PET (ROI-based) Amyloid-PET (ROI-based; APOE4 adjusted) Amyloid-PET (global) Amyloid-PET (global; APOE4 adjusted) CSF-Abeta CSF-Abeta (APOE4 adjusted) CSF-Abeta pathology (based on threshold) neuritic plaque (NP case-control) neuritic plaque (NP ordinal by CERAD)
Brain glucose uptake	FDG-PET (ROI-based) FDG-PET (ROI-based; APOE4 adjusted) FDG-PET (global) FDG-PET (global; APOE4 adjusted)
Cognition	ADAS-Cog13 ADAS-Cog13 (APOE4 adjusted) ADNI-Executive-Function ADNI-Executive-Function (APOE4 adjusted) ADNI-Memory Score ADNI-Memory Score (APOE4 adjusted) Clinical Dementia Rating - sum of boxes Cognition Cognitive Decline Mini-mental state exam RAVLT total RAVLT total (APOE4 adjusted)
Neurodegeneration	Cortex Entorhinal cortex Entorhinal cortex (APOE4 adjusted) Hippocampus Hippocampus (APOE4 adjusted)Ventricles
Neuropathology	Amyloid beta (Immunohistochemistry-based) cerebral amyloid angiopathy (CAA) Lewy body disease (LBD case-control) Lewy body disease (LBD ordinal - 3 categories) Lewy body disease (LBD ordinal - 5 categories) Global burden of pathology hippocampal sclerosis (HS) neuritic plaque (NP case-control) neuritic plaque (NP ordinal by CERAD) neurofibrillary tangle (NFT Braak ordinal I - 7 Braak stages) neurofibrillary tangle (NFT Braak ordinal II - 4 Braak groups) Tau tangles (immunohistochemistry-based) vascular brain injury (VBI case-control) vascular brain injury (VBI ordinal)

Other brain pathologies	Lewy body disease (LBD case-control)
	Lewy body disease (LBD ordinal - 3 categories)
	Lewy body disease (LBD ordinal - 5 categories)
	White matter hyperintensities
	White matter hyperintensities (APOE4 adjusted)
	hippocampal sclerosis (HS)
	vascular brain injury (VBI case-control)
	vascular brain injury (VBI ordinal)
Tau pathology	CSF-pTau
	CSF-pTau (APOE4 adjusted)
	CSF-tTau
	CSF-tTau (APOE4 adjusted)
	neurofibrillary tangle (NFT Braak ordinal I - 7 Braak stages)
	neurofibrillary tangle (NFT Braak ordinal II - 4 Braak groups)
	Tau tangles (immunohistochemistry-based)

*ADNI; **ADNI and IGAP studies

Table 8.3: Description of AD-related phenotypes with respective publications, as currently included in the AD Atlas. Unique phenotypes (n=43) are indicated in bold.

Trait	Description	Publication	Type
AD age of onset	Age at onset of AD-defined survival (AAOS) in AD cases and non-demented elderly controls. Genome-wide survival analysis was performed through Cox proportional hazards regression where the time scale is defined as age in years ('age': AAO for cases and age at last assessment for controls).	Huang et al.	GWAS
AD by proxy	Proxy phenotype for AD case-control status inferred from UK Biobank self-report data on known parental LOAD status. Includes different types of analysis including meta-analysis of maternal and paternal AD status, parental LOAD status weighted by parental age as well as meta-analysis with both proxy and clinical AD case-controls.	Marioni et al. (meta-analysis) Marioni et al. (parental) Wightman et al. Jansen et al. Bellenguez et al.	GWAS GWAS GWAS GWAS GWAS
AD case-control (clinical AD)	Clinical AD case-control. Case: met DSM-IV criteria or had a clinical dementia rating greater than zero. Control: did not meet DSM-IV criteria for dementia, had no mild cognitive impairment and - when available - a clinical dementia rating of zero.	Beecham et al.	GWAS
AD case-control (NP conservative)	Clinico-pathologic AD dementia phenotype where cases had clinical dementia with core AD neuropathologic changes, and controls were not clinically demented and had none or minimal AD neuropathologic changes. Thorough documentation of neuropathologic assessment required, including documentation of the NIA/Reagan assessment or complete documentation of both the NFT Braak stage and the NP score.	Beecham et al.	GWAS

AD case-control (NP relaxed)	Clinico-pathologic AD dementia phenotype where cases had clinical dementia with core AD neuropathologic changes, and controls were not clinically demented and had none or minimal AD neuropathologic changes. No thorough documentation of neuropathologic assessment required.	Beecham et al.	GWAS
ADAS-Cog13	Alzheimer's disease assessment scale (ADAS) - 13-item cognitive subscale.	ADNI	GWAS
ADAS-Cog13 (APOE4 adjusted)	Alzheimer's disease assessment scale (ADAS) - 13-item cognitive subscale; Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
ADNI-Executive-Function	ADNI composite score for executive function using items from different cognitive tests.	ADNI	GWAS
ADNI-Executive-Function (APOE4 adjusted)	ADNI composite score for executive function using items from different cognitive tests; Copies of APOE ε4 included as covariate.	ADNI	GWAS
ADNI-Memory Score	ADNI composite score for memory using items from different cognitive tests.	ADNI	GWAS
ADNI-Memory Score (APOE4 adjusted)	ADNI composite score for memory using items from different cognitive tests; Copies of APOE ε4 included as covariate.	ADNI	GWAS
Amyloid beta (Immunohistochemistry-based)	Immunohistochemistry-based overall Aβ load (square root).	Batra et al.	MWAS
Amyloid-PET (global; APOE4 adjusted)	Global cortical [18F] Florbetapir PET (co-registered, averaged, standardized image and voxel size, uniform resolution) SUVR (intensity-normalized using a whole cerebellum reference region); Copies of APOE ε4 included as covariate.	ADNI	GWAS
Amyloid-PET (global)	Global cortical [18F] Florbetapir PET (co-registered, averaged, standardized image and voxel size, uniform resolution) SUVR (intensity-normalized using a whole cerebellum reference region).	ADNI	GWAS

Amyloid-PET (ROI-based; APOE4 adjusted)	Global cortical [18F] Florbetapir PET (co-registered, averaged, standardized image and voxel size, uniform resolution) SUVR (intensity-normalized using a whole cerebellum reference region); SUVR value for region of interest was extracted using MarsBaR from global cortical values based on an independent comparison of ADNI-1 [11C] Pittsburgh Compound B SUVR scans (regions where AD > CN); Copies of APOE ε4 included as covariate.	ADNI	GWAS
Amyloid-PET (ROI-based)	Global cortical [18F] Florbetapir PET (co-registered, averaged, standardized image and voxel size, uniform resolution) SUVR (intensity-normalized using a whole cerebellum reference region); SUVR value for region of interest was extracted using MarsBaR from global cortical values based on an independent comparison of ADNI-1 [11C] Pittsburgh Compound B SUVR scans (regions where AD > CN).	ADNI	GWAS
APOE4	Metabolite associations with copies of APOE ε4 (additive model)	Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	MWAS MWAS
cerebral amyloid angiopathy (CAA)	Cerebral amyloid angiopathy (CAA) analyzed as co-morbid neuropathologic phenotype using presence vs. absence analyses.	Beecham et al.	GWAS
Clinical Dementia Rating - sum of boxes	Copies of APOE ε4 included as covariate.	Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	MWAS MWAS
Clinical diagnosis (ordinal)	Consensus cognitive diagnosis at time of death analyzed with ordinal ranking - three categories: AD, Mild cognitive impairment (MCI), No cognitive impairment (NCI).	Batra et al.	MWAS

CN vs. AD	Case-control AD study. Includes meta-analysis.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al. Lambert et al. Kunkle et al.	GWAS MWAS MWAS GWAS GWAS
CN vs. MCI	Case-control study of CN participants vs. participants with MCI (early and late).	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
Cognition	Global cognitive function determined at the last timepoint before death.	Batra et al.	MWAS
Cognitive decline	Rate of change in global cognition over time.	Batra et al.	MWAS
Composite measure of brain atrophy (cross-region MRI analysis)	Composite measure of atrophy in regions affected by AD (SPARE-AD).	Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	MWAS MWAS
Cortex	Global cortical grey matter volume based on MRI. Adjusted for Copies of APOE ε4.	Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	MWAS MWAS
CSF-Abeta	Amyloid beta (Aβ1-42) levels measured in cerebrospinal fluid (CSF). For meta-analysis the raw values were log10-transformed to approximate a normal distribution within each study and centralized by each study mean.	ADNI Deming et al.	GWAS GWAS
CSF-Abeta (APOE4 adjusted)	Amyloid beta (Aβ1-42) levels measured in cerebrospinal fluid (CSF) using the Roche Elecsys immunoassay. Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
CSF-Abeta pathology (based on threshold)	Amyloid beta (Aβ1-42) positivity in cerebrospinal fluid (CSF) based on measures using the Roche Elecsys immunoassay and a threshold of 1073 pg/ml (positive; = negative); Copies of APOE ϵ4 included as covariate.	Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	MWAS MWAS

CSF-CLU	Clusterin (CLU) levels measured in cerebrospinal fluid (CSF). For meta-analysis the raw values were log10-transformed to approximate a normal distribution within each study and centralized by each study mean.	Deming et al.	GWAS
CSF-pTau	Phosphorylated tau (ptau181) levels measured in cerebrospinal fluid (CSF). For meta-analysis the raw values were log10-transformed to approximate a normal distribution within each study and centralized by each study mean.	ADNI Deming et al.	GWAS GWAS
CSF-pTau (APOE4 adjusted)	Phosphorylated tau (ptau181) levels measured in cerebrospinal fluid (CSF) using the Roche Elecsys immunoassay. Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
CSF-pTau/Abeta	Ratio of phosphorylated tau levels and amyloid beta 1-42 levels measured in cerebrospinal fluid (CSF) using the Roche Elecsys immunoassay.	ADNI	GWAS
CSF-pTau/Abeta (APOE4 adjusted)	Ratio of phosphorylated tau levels and amyloid beta 1-42 levels measured in cerebrospinal fluid (CSF) using the Roche Elecsys immunoassay; Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
CSF-tTau	Total tau levels measured in cerebrospinal fluid (CSF) using the Roche Elecsys immunoassay. For meta-analysis the raw values were log10-transformed to approximate a normal distribution within each study and centralized by each study mean.	ADNI Deming et al.	GWAS GWAS
CSF-tTau (APOE4 adjusted)	Total tau levels measured in cerebrospinal fluid (CSF) using the Roche Elecsys immunoassay; Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
CSF-tTau/Abeta	Ratio of total tau levels and amyloid beta 1-42 levels measured in cerebrospinal fluid (CSF) using the Roche Elecsys immunoassay.	ADNI	GWAS

CSF-tTau/Abeta (APOE4 adjusted)	Ratio of total tau levels and amyloid beta 1-42 levels measured in cerebrospinal fluid (CSF) using the Roche Elecsys immunoassay; Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
Entorhinal cortex	Entorhinal cortical thickness from MRI.	ADNI	GWAS
Entorhinal cortex (APOE4 adjusted)	Entorhinal cortical thickness from MRI; Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
FDG-PET (global; APOE4 adjusted)	Global cortical [18F] FDG PET (coregistered, averaged, standardized image and voxel size, uniform resolution) SUVR (intensity-normalized using a pons reference region); Copies of APOE ε4 included as covariate.	ADNI	GWAS
FDG-PET (global)	Global cortical [18F] FDG PET (coregistered, averaged, standardized image and voxel size, uniform resolution) SUVR (intensity-normalized using a pons reference region).	ADNI	GWAS
FDG-PET (ROI-based; APOE4 adjusted)	Global cortical [18F] FDG PET (coregistered, averaged, standardized image and voxel size, uniform resolution) SUVR (intensity-normalized using a pons reference region); a mean SUVR value was extracted from global cortical values representing regions where AD patients show decreased glucose metabolism relative to cognitively normal older participants (CN) from the full ADNI-1 cohort; Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS

FDG-PET (ROI-based)	Global cortical [18F] FDG PET (co-registered, averaged, standardized image and voxel size, uniform resolution) SUVR (intensity-normalized using a pons reference region); a mean SUVR value was extracted from global cortical values representing regions where AD patients show decreased glucose metabolism relative to cognitively normal older participants (CN) from the full ADNI-1 cohort.	ADNI	GWAS
Global burden of pathology	Summary of pathology derived from counts of: neuritic plaques, diffuse plaques, and neurofibrillary tangles.	Batra et al.	MWAS
hippocampal sclerosis (HS)	Hippocampal sclerosis (HS) analyzed as co-morbid neuropathologic phenotype using presence vs. absence analyses.	Beecham et al.	GWAS
Hippocampus	Hippocampal grey matter volume from MRI.	ADNI	GWAS
Hippocampus (APOE4 adjusted)	Hippocampal grey matter volume from MRI; Copies of APOE ε4 included as covariate.	ADNI Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	GWAS MWAS MWAS
Lewy body disease (LBD case-control)	Lewy body disease (LBD) analyzed with a presence vs. absence - any LBD vs. no LBD - analysis.	Beecham et al.	GWAS
Lewy body disease (LBD ordinal - 3 categories)	Lewy body disease (LBD) analyzed with ordinal ranking - three categories: none, brainstem-predominant, and all other regions or not specified.	Beecham et al.	GWAS
Lewy body disease (LBD ordinal - 5 categories)	Lewy body disease (LBD) analyzed with ordinal ranking - five categories: none, brainstem-predominant, limbic, neocortical, and other regions or not specified.	Beecham et al.	GWAS
MCI vs. AD	case/control study of ADNI participants with MCI (early and late) vs. cases with clinical AD.	ADNI	GWAS
Mini-mental state exam	Adjusted for Copies of APOE ε4.	Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	MWAS MWAS

neuritic plaque (NP case-control)	Neuritic plaques (NPs) analyzed with a presence vs. absence - any NPs vs. no NPs - analysis.	Beecham et al.	GWAS
neuritic plaque (NP ordinal by CERAD)	Neuritic plaques (NPs) analyzed with ordinal ranking - four CERAD scores: none, sparse, moderate, frequent.	Beecham et al.	GWAS
neurofibrillary tangle (NFT Braak ordinal I - 7 Braak stages)	Neurofibrillary tangles (NFTs) analyzed by well-established ordinal ranking - seven Braak stages: none, I, II, III, IV, V, and VI.	Beecham et al.	GWAS
neurofibrillary tangle (NFT Braak ordinal II - 4 Braak groups)	Neurofibrillary tangles (NFTs) analyzed by well-established ordinal ranking - four Braak groups: none, transentorhinal, limbic, isocortical.	Beecham et al.	GWAS
Neuropathology diagnosis	Mayo clinic post-mortem diagnosis of Alzheimer's disease derived from Braak and CERAD scores. AD case status was assigned where Braak stage was ≥ 4 and CERAD score was ≤ 2; control case status was assigned where Braak stage was ≤ 3 and CERAD score was ≥ 3.	Batra et al.	MWAS
NIA-Reagan Score	NIA Reagan diagnosis of Alzheimer's disease derived from Braak and CERAD scores; binarized (0: low likelihood of AD, 1: high likelihood of AD).	Batra et al.	MWAS
No AD vs. AD	Case/control setting combining all non-demented individuals (CN, SMC, EMCI, LMCI) into the control group vs. cases with clinical AD.	ADNI	GWAS
Progression MCI -> AD	Ever-progression (coded as 0/1) observed in ADNI participants during follow-up (up to ten years) with baseline diagnosis of late MCI.	Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	MWAS MWAS
RAVLT total	Rey Auditory-Verbal Learning Test (RAVLT).	ADNI	GWAS
RAVLT total (APOE4 adjusted)	Rey Auditory-Verbal Learning Test (RAVLT); Copies of APOE ε4 included as covariate.	ADNI	GWAS
Tau tangles (immunohistochemistry-based)	Immunohistochemistry-based overall paired helical filament (PHF)-tau tangles load.	Batra et al.	MWAS

vascular brain injury (VBI case-control)	Vascular brain injury (VBI) analyzed with a presence vs. absence - any VBI vs. no VBI - analysis.	Beecham et al.	GWAS
vascular brain injury (VBI ordinal)	Vascular brain injury (VBI) analyzed with ordinal ranking - three categories: none, any microinfarcts, any lacunar or territorial infarcts.	Beecham et al.	GWAS
Ventricles	Ventricular grey matter volume from MRI.	Arnold et al. Mahmoudian-Dehkordi et al. and Nho et al.	MWAS MWAS
White matter hyperintensities	Global cortical white matter hyperintensities from MRI.	ADNI	GWAS
White matter hyperintensities (APOE4 adjusted)	Global cortical white matter hyperintensities from MRI; Copies of APOE ε4 included as covariate.	ADNI	GWAS

Table 8.4: Number of measured metabolites that are consolidated in each meta metabolite listed with the number of their respective occurrence. For example, n=148 meta metabolites consolidate three measured metabolites. An example is given in Figure 4.4.

Number of measured metabolites/compounds consolidated to one "meta Metabolite"	**n**
1 = no consolidation	806
2	318
3	148
4	37
5	12
6	6
7	1

Table 8.5: Details on the metabolomics platform providers for the 522 meta metabolites that consolidate two or more measured metabolites. For example, cross-platform mapping was performed for n=60 metabolites.

Metabolomics platform provider	**n**
Metabolon	462
Biocrates	0
both	60
Total meta metabolites that consolidate >1 measured metabolite:	522

Table 8.6: Clusters with significant enrichment for biodomains. Clusters were obtained by applying hierarchical clustering on the 89 dimensional (D) node vectors and using cut h = 30. Multiple testing correction was performed using Bonferroni, resulting in a significance threshold of p-value $\leq$ 0.00217 (0.05/23 clusters).

Cluster*	**Biodomain**	**p-value**	$\mathbf{p}_{adj}$	**OR**	**95% CI****
1	Myelination	1.49e-04	3.43e-03	2.59	1.71
2	Apoptosis	6.15e-10	1.42e-08	2.30	1.84
2	Immune Response	8.17e-15	1.88e-13	2.56	2.10
2	Oxidative Stress	5.49e-04	1.26e-02	2.07	1.45
2	Vasculature	1.37e-13	3.16e-12	2.94	2.33
3	Myelination	3.16e-09	7.26e-08	3.30	2.40
4	APP Metabolism	1.01e-11	2.33e-10	9.86	6.05
4	Proteostasis	1.27e-10	2.92e-09	3.52	2.46
4	Structural Stabilization	2.98e-14	6.85e-13	4.33	3.05
4	Synapse	5.13e-04	1.18e-02	2.10	1.45
5	Autophagy	1.11e-12	2.54e-11	3.24	2.50
5	Endolysosome	7.70e-08	1.77e-06	2.12	1.69
5	Immune Response	1.78e-64	4.09e-63	7.35	6.01
5	Lipid Metabolism	1.42e-08	3.26e-07	1.99	1.63
6	Synapse	7.62e-50	1.75e-48	2.66	2.39
7	Vasculature	4.89e-06	1.13e-04	1.60	1.35
8	Mitochondrial Metabolism	1.84e-59	4.23e-58	4.14	3.60
9	RNA Spliceosome	8.13e-14	1.87e-12	4.13	3.09
10	Structural Stabilization	4.83e-10	1.11e-08	1.96	1.63
10	Vasculature	4.70e-29	1.08e-27	4.42	3.59
11	RNA Spliceosome	3.50e-05	8.06e-04	1.67	1.35
12	Immune Response	8.06e-25	1.85e-23	2.47	2.14
12	Lipid Metabolism	2.08e-03	4.79e-02	1.30	1.12
12	Vasculature	2.28e-05	5.24e-04	1.64	1.35

*corresponding to clusters indicated in Figure 5.3 of the main manuscript; **lower bound of one-sided test reported

Table 8.7: Statins listed in DrugBank and their annotated targets.

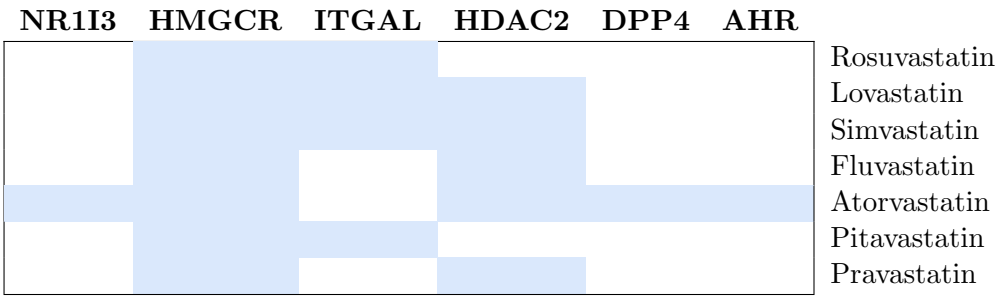


Table 8.8: Genes co-expressed with *TMEM119* but not *TREM2* that show genetic associations with AD-related phenotypes. The lowest p-value is reported for each locus.

Gene symbol	**AD-related phenotype**	**p-value**
ARHGAP45	AD by proxy	4.09e-30
ARHGAP45	CN vs. AD	3.06e-16
ARPC1B	AD by proxy	4.83e-11
ATP8B4	AD by proxy	5.12e-10
GPSM3	AD by proxy	8.76e-11
HLA-DMA	AD by proxy	2.09e-16
HLA-DMA	ADAS-Cog13	1.04e-09
HLA-DMA	ADAS-Cog13 (APOE4 adjusted)	3.15e-09
HLA-DMA	ADNI-Executive-Function	4.81e-08
HLA-DMA	ADNI-Memory Score	3.09e-08
HLA-DMA	CN vs. AD	5.10e-12
HLA-DMA	MCI vs. AD	3.16e-10
HLA-DMA	No AD vs. AD	2.78e-10
INPP5D	AD by proxy	1.04e-17
INPP5D	CN vs. AD	3.42e-09
SPN	AD by proxy	5.39e-10
TMEM106A	ADAS-Cog13	2.95e-10
TMEM106A	ADAS-Cog13 (APOE4 adjusted)	4.92e-09
TMEM106A	ADNI-Memory Score	3.52e-09
TMEM106A	CN vs. AD	2.56e-08

Table 8.9: Genes co-expressed with *TMEM119* but not *TREM2* that show genetic associations with AD phenotypes and metabolite traits (metabolite quantitative trait loci (mQTL)). The lowest p-value is reported for each locus.

Gene symbol	**Metabolic trait**	**p-value**
ARPC1B	16a-hydroxy DHEA 3-sulfate	1.46e-28
ARPC1B	4-androsten-3alpha,17alpha-diol monosulfate (3)	1.85e-27
ARPC1B	4-androsten-3beta,17beta-diol disulfate (1)	6.27e-09
ARPC1B	4-androsten-3beta,17beta-diol monosulfate (1)	1.52e-11
ARPC1B	5alpha-androstan-3alpha,17beta-diol monosulfate (1)	2.97e-09
ARPC1B	5alpha-androstan-3beta,17beta-diol disulfate	9.27e-33
ARPC1B	andro steroid monosulfate (1)*	1.40e-21
ARPC1B	androsterone sulfate	8.82e-113
ARPC1B	dehydroisoandrosterone sulfate (DHEA-S)	6.70e-14
ARPC1B	epiandrosterone sulfate	2.80e-75
ARPC1B	X-12063**	1.67e-109
INPP5D	1-arachidonoyl-GPA (20:4)	1.68e-09
INPP5D	bilirubin	3.21e-11
INPP5D	biliverdin	6.33e-17
INPP5D	X-11441	5.48e-17
INPP5D	X-11442	1.56e-16
INPP5D	X-11530	6.42e-13

**metabolonic lactone sulfate (partially characterized metabolite)

Table 8.10: Genes co-expressed with *TREM2* but not *TMEM119* that show genetic associations with AD-related phenotypes. The lowest p-value is reported for each locus.

Gene symbol	AD-related phenotype	p-value
AIF1	AD by proxy	1.19e-16
AIF1	CN vs. AD	7.16e-09
APOC1	AD age of onset	4.32e-131
APOC1	AD by proxy	0.00e+00
APOC1	AD case-control (clinical AD)	2.64e-63
APOC1	AD case-control (NP conservative)	3.48e-38
APOC1	AD case-control (NP relaxed)	2.02e-62
APOC1	ADAS-Cog13	1.04e-29
APOC1	ADNI-Executive-Function	2.80e-17
APOC1	ADNI-Memory Score	4.29e-38
APOC1	Amyloid-PET (global)	1.14e-44
APOC1	Amyloid-PET (ROI-based)	1.16e-42
APOC1	cerebral amyloid angiopathy (CAA)	2.92e-21
APOC1	CN vs. AD	0.00e+00
APOC1	CN vs. MCI	4.30e-15
APOC1	CSF-Abeta	4.78e-94
APOC1	CSF-Abeta (APOE4 adjusted)	4.67e-16
APOC1	CSF-pTau	5.67e-37
APOC1	CSF-pTau/Abeta	8.84e-95
APOC1	CSF-pTau/Abeta (APOE4 adjusted)	3.17e-16
APOC1	CSF-tTau	4.61e-30
APOC1	CSF-tTau/Abeta	7.97e-97
APOC1	CSF-tTau/Abeta (APOE4 adjusted)	1.76e-17
APOC1	Entorhinal cortex	1.14e-10
APOC1	FDG-PET (global)	7.15e-10
APOC1	FDG-PET (ROI-based)	9.45e-16
APOC1	Hippocampus	1.80e-27
APOC1	Lewy body disease (LBD case-control)	2.83e-11
APOC1	Lewy body disease (LBD ordinal - 3 categories)	4.87e-12
APOC1	Lewy body disease (LBD ordinal - 5 categories)	1.10e-12
APOC1	MCI vs. AD	8.24e-09
APOC1	neuritic plaque (NP case-control)	1.78e-27
APOC1	neuritic plaque (NP ordinal by CERAD)	1.37e-46
APOC1	neurofibrillary tangle (NFT Braak ordinal I - 7 Braak stages)	4.73e-47
APOC1	neurofibrillary tangle (NFT Braak ordinal II - 4 Braak groups)	4.83e-44

APOC1	No AD vs. AD	3.66e-18
APOC1	RAVLT total	1.98e-26
GAL3ST4	AD by proxy	9.92e-19
GAL3ST4	CN vs. AD	5.58e-10
HLA-DPB1	AD by proxy	4.34e-16
HLA-DPB1	ADAS-Cog13	2.68e-11
HLA-DPB1	ADAS-Cog13 (APOE4 adjusted)	6.17e-11
HLA-DPB1	ADNI-Executive-Function	4.19e-09
HLA-DPB1	ADNI-Executive-Function (APOE4 adjusted)	1.30e-08
HLA-DPB1	ADNI-Memory Score	3.18e-10
HLA-DPB1	ADNI-Memory Score (APOE4 adjusted)	8.76e-10
HLA-DPB1	CN vs. AD	2.94e-12
HLA-DPB1	MCI vs. AD	2.25e-12
HLA-DPB1	No AD vs. AD	1.62e-12
HLA-DRB1	AD by proxy	1.05e-17
HLA-DRB1	ADAS-Cog13	6.16e-09
HLA-DRB1	ADNI-Memory Score	1.00e-09
HLA-DRB1	ADNI-Memory Score (APOE4 adjusted)	8.24e-09
HLA-DRB1	CN vs. AD	2.94e-12
HLA-DRB1	Entorhinal cortex	6.18e-09
HLA-DRB1	MCI vs. AD	6.42e-11
HLA-DRB1	No AD vs. AD	3.00e-11
HLA-DRB1	RAVLT total	3.52e-08
ITGAM	AD by proxy	8.50e-10
ITGAX	AD by proxy	7.71e-10
LST1	AD by proxy	1.19e-16
LST1	CN vs. AD	1.38e-09
SPI1	AD by proxy	7.84e-12
SPI1	CN vs. AD	5.46e-13

Table 8.11: Genes co-expressed with *TREM2* but not *TMEM119* that show genetic associations with AD phenotypes and metabolite traits (mQTLs). The lowest p-value is reported for each locus.

Gene symbol	**Metabolic trait**	**p-value**
AIF1	X-12712	2.12e-12
APOC1	cholesterol	4.10e-10
APOC1	SM (OH) C22:1	1.36e-12
APOC1	SM (OH) C22:2	2.23e-08
APOC1	SM (OH) C24:1	7.91e-11
APOC1	SM C16:0	1.97e-09
APOC1	SM C18:0	1.16e-08
APOC1	SM C24:0	9.31e-10
APOC1	X-11820	1.22e-16
GAL3ST4	16a-hydroxy DHEA 3-sulfate	9.09e-12
GAL3ST4	4-androsten-3alpha,17alpha-diol monosulfate (3)	1.65e-13
GAL3ST4	5alpha-androstan-3beta,17beta-diol disulfate	1.76e-14
GAL3ST4	andro steroid monosulfate (1)*	4.91e-10
GAL3ST4	androsterone sulfate	3.07e-31
GAL3ST4	epiandrosterone sulfate	9.11e-24
GAL3ST4	X-12063**	1.67e-109
HLA-DPB1	N6-methyllysine	3.87e-08
HLA-DRB1	X-11470	6.56e-13
LST1	X-11470	3.41e-09

**metabolonic lactone sulfate (partially characterized metabolite)

List of Abbreviations

Aβ	amyloid-β
ACT	Adult Changes in Though
AD	Alzheimer's disease
ADAS-Cog	Alzheimer's Disease Assessment Scale-Cognitive Subscale
ADC	Alzheimer's Disease Center
ADGC	Alzheimer's Disease Genetics Consortium
ADMC	Alzheimer's Disease Metabolomics Consortium
ADNI	Alzheimer's Disease Neuroimaging Initiative
ADRD	Alzheimer's disease and related dementia's
ADSP	Alzheimer's Disease Sequencing Project
AMP-AD	Accelerating Medicines Partnership - Alzheimer's Disease
AOD	age of death
APOE	apolipoprotein E
APP	amyloid beta precursor protein
BACE-1	beta-site amyloid precursor protein–cleaving enzyme 1
BLSA	Baltimore Longitudinal Study of Aging
Banner	Banner Sun Health Research Institute
BCAA	branched-chain amino acid
BCKD	branched-chain alpha-ketoacid dehydrogenase
BIOCARD	Predictors of Cognitive Decline Among Normal Individuals
BMI	body mass index
CBE	cerebellum
CBMM	constraint-based metabolic model
CDR	Clinical Dementia Rating
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CLU	cerebrospinal fluid clusterin
CMap	Broad Institute Connectivity Map
CNS	central nervous system
COBRA	Constraint-Based Reconstruction and Analysis
COMP ID	compound identifier
CSF	cerebrospinal fluid
CSV	comma-separated values
CyTOF	cytometry time of flight
D	dimensional
DAM	disease-associated microglia
DB	database
DEG	differentially expressed gene
DEP	differentially abundant protein
DIABLO	Data Integration Analysis for Biomarker discovery using Latent cOmponents
DLPFC	dorsolateral prefrontal cortex
DNA	deoxyribonucleic acid
DR	dimensionality reduction

EADB	European Alzheimer & Dementia Biobank
EADI	European Alzheimer's Disease Initiative
EOAD	early-onset Alzheimer's disease
eQTL	expression quantitative trait loci
FDA	Food and Drug Administration
FDG	fluorodeoxyglucose
FDR	False Discovery Rate
FP	frontal pole
FSEA	Functional set enrichment analysis
GEM	genome-wide metabolic model
GEO	Gene Expression Omnibus
GERAD	Genetic and Environmental Risk in Alzheimer's Disease
GGM	Gaussian Graphical Models
GO	Gene Ontology
GSEA	Gene set enrichment analysis
GTEx	Genotype-Tissue Expression
GWAS	genome-wide association studies
HLA	human leukocyte antigen
HMGCR	3-hydroxy-3-methylglutaryl coenzyme A reductase
hQTL	histone quantitative trait loci
HRC	Haplotype Reference Consortium
ID	identifier
IDA	Image & Data Archive
IFG	inferior frontal gyrus
IGAP	International Genomics of Alzheimer's Project
IMPaLA	Integrated Molecular Pathway-Level Analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
Knight ADRC	Charles F. and Joanne Knight Alzheimer's Disease Research Center
LFA-1	integrin leukocyte function-associated antigen-1
LFQ-MS	label-free quantitation mass spectrometry
LMSD	LIPID MAPS Structure Database
LOAD	late-onset Alzheimer's disease
LONI	Laboratory of Neuroimaging
MAF	minor allele frequency
MALDI	Matrix Assisted Laser Desorption Ionization
MAP	Memory and Aging Project
Mayo	Mayo Clinic Brain Bank
MCI	Mild Cognitive Impairment
meQTL	methylation quantitative trait loci
MGM	Mixed Graphical Models
mGWAS	genome-wide association studies with metabolite traits
MOFA	Multi-Omics Factor Analysis
mQTL	metabolite quantitative trait loci
MS	mass spectrometry
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MSBB	Mount Sinai Brain Bank
MWAS	metabolome-wide association studies
NDEx	Network Data Exchange
NIA	National Institutes on Aging
NIAGADS	NIA Genetics of Alzheimer's Disease Data Storage Site
NIH	National Institutes of Health
NMDA	*N*-methyl-D-aspartate
OmicsDI	Omics Discovery Index
OnPLS	Multiple-Block Orthogonal Projections to Latent Structures

ORA	over-representation analysis
PC	phosphatidylcholines
PCA	Principal Component Analysis
PET	positron emission tomography
PHG	parahippocampal gyrus
p-tau	phosphorylated tau
PGC-ALZ	Alzheimer's disease working group of the Psychiatric Genomics Consortium
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
PMI	post mortem interval
PPI	protein-protein interaction
pQTL	protein quantitative trait loci
PSEN1	presenilin 1
PSEN2	presenilin 2
QC	quality control
QTL	quantitative trait loci
RGCCA	regularized generalized canonical correlation analysis
ROCK	Rho-kinase
RNA	ribonucleic acid
ROS	Religious Orders Study
rsID	reference SNP cluster ID
S1P	sphingosine-1-phosphate
SGCCA	sparse generalized canonical correlation
SM	sphingomyelin
SNP	single nucleotide polymorphism
sQTL	splicing quantitative trait loci
SSRI	selective serotonin reuptake inhibitor
STG	superior temporal gyrus
TCX	temporal cortex
TMT-MS	tandem mass tag mass spectrometry
TNFα	tumor necrosis factor α
t-tau	total tau
UKBB	UK Biobank
UMAP	Uniform Manifold Approximation and Projection
VAE	variational autoencoder
WGCNA	weighted gene co-expression network analysis
WGS	whole-genome sequencing

List of Figures

List of Tables

Bibliography

[1] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," Genome Biol., vol. 18, p. 83, May 2017.

[2] F. R. Pinu, D. J. Beale, A. M. Paten, K. Kouremenos, S. Swarup, H. J. Schirra, and D. Wishart, "Systems biology and Multi-Omics integration: Viewpoints from the metabolomics research community," Metabolites, vol. 9, pp. 1–31, Apr. 2019.

[3] D. P. Veitch, M. W. Weiner, P. S. Aisen, L. A. Beckett, C. DeCarli, R. C. Green, D. Harvey, C. R. Jack, Jr, W. Jagust, S. M. Landau, J. C. Morris, O. Okonkwo, R. J. Perrin, R. C. Petersen, M. Rivera-Mindt, A. J. Saykin, L. M. Shaw, A. W. Toga, D. Tosun, J. Q. Trojanowski, and Alzheimer's Disease Neuroimaging Initiative, "Using the alzheimer's disease neuroimaging initiative to improve early detection, diagnosis, and treatment of alzheimer's disease," Alzheimers. Dement., vol. 18, pp. 824–857, Apr. 2022.

[4] R. J. Hodes and N. Buckholtz, "Accelerating medicines partnership: Alzheimer's disease (AMP-AD) knowledge portal aids alzheimer's drug discovery through open data sharing," Expert Opin. Ther. Targets, vol. 20, pp. 389–391, Feb. 2016.

[5] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, and R. Collins, "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age," PLoS Med., vol. 12, p. e1001779, Mar. 2015.

[6] B. B. Misra, C. D. Langefeld, M. Olivier, and L. A. Cox, "Integrated omics: Tools, advances, and future approaches," J. Mol. Endocrinol., pp. R21–R45, July 2018.

[7] A. Lysenko, I. A. Roznovăţ, M. Saqi, A. Mazein, C. J. Rawlings, and C. Auffray, "Representing and querying disease networks using graph databases," Dec. 2016.

[8] M. Koutrouli, E. Karatzas, D. Paez-Espino, and G. A. Pavlopoulos, "A guide to conquer the biological network era using graph theory," Front Bioeng Biotechnol, vol. 8, p. 34, Jan. 2020.

[9] S. Choobdar, M. E. Ahsen, J. Crawford, M. Tomasoni, T. Fang, D. Lamparter, J. Lin, B. Hescott, X. Hu, J. Mercer, T. Natoli, R. Narayan, DREAM Module Identification Challenge Consortium, A. Subramanian, J. D. Zhang, G. Stolovitzky, Z. Kutalik, K. Lage, D. K. Slonim, J. Saez-Rodriguez, L. J. Cowen, S. Bergmann, and D. Marbach, "Assessment of network module identification across complex diseases," Nat. Methods, vol. 16, pp. 843–852, Sept. 2019.

[10] M. A. Wörheide, J. Krumsiek, G. Kastenmüller, and M. Arnold, "Multi-omics integration in biomedical research - a metabolomics-centric review," Anal. Chim. Acta, vol. 1141, pp. 144–162, Jan. 2021.

[11] L. A. Lotta, R. A. Scott, S. J. Sharp, S. Burgess, J. Luan, T. Tillin, A. F. Schmidt, F. Imamura, I. D. Stewart, J. R. B. Perry, L. Marney, A. Koulman, E. D. Karoly, N. G. Forouhi, R. J. O. Sjögren, E. Näslund, J. R. Zierath, A. Krook, D. B. Savage, J. L. Griffin, N. Chaturvedi, A. D. Hingorani,

K.-T. Khaw, I. Barroso, M. I. McCarthy, S. O'Rahilly, N. J. Wareham, and C. Langenberg, "Genetic predisposition to an impaired metabolism of the Branched-Chain amino acids and risk of type 2 diabetes: A mendelian randomisation analysis," PLoS Med., vol. 13, p. e1002179, Nov. 2016.

[12] J. Tynkkynen, V. Chouraki, S. J. van der Lee, J. Hernesniemi, Q. Yang, S. Li, A. Beiser, M. G. Larson, K. Sääksjärvi, M. J. Shipley, A. Singh-Manoux, R. E. Gerszten, T. J. Wang, A. S. Havulinna, P. Würtz, K. Fischer, A. Demirkan, M. A. Ikram, N. Amin, T. Lehtimäki, M. Kähönen, M. Perola, A. Metspalu, A. J. Kangas, P. Soininen, M. Ala-Korpela, R. S. Vasan, M. Kivimäki, C. M. van Duijn, S. Seshadri, and V. Salomaa, "Association of branched-chain amino acids and other circulating metabolites with risk of incident dementia and alzheimer's disease: A prospective study in eight cohorts," Alzheimers. Dement., vol. 14, pp. 723–733, June 2018.

[13] M. Civelek and A. J. Lusis, "Systems genetics approaches to understand complex traits," Nat. Rev. Genet., vol. 15, pp. 34–48, Jan. 2014.

[14] G. D. Smith and S. Ebrahim, "'mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?," Int. J. Epidemiol., vol. 32, pp. 1–22, Feb. 2003.

[15] D. Kopczynski, C. Coman, R. P. Zahedi, K. Lorenz, A. Sickmann, and R. Ahrends, "Multi-OMICS: a critical technical perspective on integrative lipidomics approaches," Biochim. Biophys. Acta Mol. Cell Biol. Lipids, vol. 1862, pp. 808–811, Aug. 2017.

[16] M. D. Ritchie, E. R. Holzinger, R. Li, S. A. Pendergrass, and D. Kim, "Methods of integrating data to uncover genotype-phenotype interactions," Nat. Rev. Genet., vol. 16, pp. 85–97, Feb. 2015.

[17] D. S. Wishart, "Emerging applications of metabolomics in drug discovery and precision medicine," Nat. Rev. Drug Discov., vol. 15, no. 7, pp. 473–484, 2016.

[18] O. Fiehn, "Metabolomics–the link between genotypes and phenotypes," Plant Mol. Biol., vol. 48, pp. 155–171, Jan. 2002.

[19] J. B. Toledo, M. Arnold, G. Kastenmüller, R. Chang, R. A. Baillie, X. Han, M. Thambisetty, J. D. Tenenbaum, K. Suhre, J. W. Thompson, L. S. John-Williams, S. MahmoudianDehkordi, D. M. Rotroff, J. R. Jack, A. Motsinger-Reif, S. L. Risacher, C. Blach, J. E. Lucas, T. Massaro, G. Louie, H. Zhu, G. Dallmann, K. Klavins, T. Koal, S. Kim, K. Nho, L. Shen, R. Casanova, S. Varma, C. Legido-Quigley, M. A. Moseley, K. Zhu, M. Y. R. Henrion, S. J. van der Lee, A. C. Harms, A. Demirkan, T. Hankemeier, C. M. van Duijn, J. Q. Trojanowski, L. M. Shaw, A. J. Saykin, M. W. Weiner, P. M. Doraiswamy, R. Kaddurah-Daouk, and Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium, "Metabolic network failures in alzheimer's disease: A biochemical road map," Alzheimers. Dement., vol. 13, pp. 965–984, Sept. 2017.

[20] K. Suhre, C. Meisinger, A. Döring, E. Altmaier, P. Belcredi, C. Gieger, D. Chang, M. V. Milburn, W. E. Gall, K. M. Weinberger, H.-W. Mewes, M. Hrabé de Angelis, H.-E. Wichmann, F. Kronenberg, J. Adamski, and T. Illig, "Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting," PLoS One, vol. 5, p. e13953, Nov. 2010.

[21] M. Yang, T. Soga, and P. J. Pollard, "Oncometabolites: linking altered metabolism with cancer," J. Clin. Invest., vol. 123, pp. 3652–3658, Sept. 2013.

[22] R. D. Beger, W. Dunn, M. A. Schmidt, S. S. Gross, J. A. Kirwan, M. Cascante, L. Brennan, D. S. Wishart, M. Oresic, T. Hankemeier, D. I. Broadhurst, A. N. Lane, K. Suhre, G. Kastenmüller, S. J. Sumner, I. Thiele, O. Fiehn, R. Kaddurah-Daouk, and for "Precision Medicine and Pharmacometabolomics Task Group"-Metabolomics Society Initiative, "Metabolomics enables precision medicine: "a white paper, community perspective"," Metabolomics, vol. 12, p. 149, Sept. 2016.

[23] C. Gieger, L. Geistlinger, E. Altmaier, M. Hrabé de Angelis, F. Kronenberg, T. Meitinger, H.-W. Mewes, H.-E. Wichmann, K. M. Weinberger, J. Adamski, T. Illig, and K. Suhre, "Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum," PLoS Genet., vol. 4, p. e1000282, Nov. 2008.

[24] H. H. M. Draisma, R. Pool, M. Kobl, R. Jansen, A.-K. Petersen, A. A. M. Vaarhorst, I. Yet, T. Haller, A. Demirkan, T. Esko, G. Zhu, S. Böhringer, M. Beekman, J. B. van Klinken, W. Römisch-Margl, C. Prehn, J. Adamski, A. J. M. de Craen, E. M. van Leeuwen, N. Amin, H. Dharuri, H.-J. Westra, L. Franke, E. J. C. de Geus, J. J. Hottenga, G. Willemsen, A. K. Henders, G. W. Montgomery, D. R. Nyholt, J. B. Whitfield, B. W. Penninx, T. D. Spector, A. Metspalu, P. E. Slagboom, K. W. van Dijk, P. A. C. 't Hoen, K. Strauch, N. G. Martin, G.-J. B. van Ommen, T. Illig, J. T. Bell, M. Mangino, K. Suhre, M. I. McCarthy, C. Gieger, A. Isaacs, C. M. van Duijn, and D. I. Boomsma, "Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels," Nat. Commun., vol. 6, p. 7208, June 2015.

[25] S.-Y. Shin, E. B. Fauman, A.-K. Petersen, J. Krumsiek, R. Santos, J. Huang, M. Arnold, I. Erte, V. Forgetta, T.-P. Yang, K. Walter, C. Menni, L. Chen, L. Vasquez, A. M. Valdes, C. L. Hyde, V. Wang, D. Ziemek, P. Roberts, L. Xi, E. Grundberg, Multiple Tissue Human Expression Resource (MuTHER) Consortium, M. Waldenberger, J. B. Richards, R. P. Mohney, M. V. Milburn, S. L. John, J. Trimmer, F. J. Theis, J. P. Overington, K. Suhre, M. J. Brosnan, C. Gieger, G. Kastenmüller, T. D. Spector, and N. Soranzo, "An atlas of genetic influences on human blood metabolites," Nat. Genet., vol. 46, pp. 543–550, June 2014.

[26] J. Raffler, N. Friedrich, M. Arnold, T. Kacprowski, R. Rueedi, E. Altmaier, S. Bergmann, K. Budde, C. Gieger, G. Homuth, M. Pietzner, W. Römisch-Margl, K. Strauch, H. Völzke, M. Waldenberger, H. Wallaschofski, M. Nauck, U. Völker, G. Kastenmüller, and K. Suhre, "Genome-Wide association study with targeted and non-targeted NMR metabolomics identifies 15 novel loci of urinary human metabolic individuality," PLoS Genet., vol. 11, p. e1005487, Sept. 2015.

[27] J. Bartel, J. Krumsiek, K. Schramm, J. Adamski, C. Gieger, C. Herder, M. Carstensen, A. Peters, W. Rathmann, M. Roden, K. Strauch, K. Suhre, G. Kastenmüller, H. Prokisch, and F. J. Theis, "The human blood Metabolome-Transcriptome interface," PLoS Genet., vol. 11, p. e1005274, June 2015.

[28] A.-K. Petersen, S. Zeilinger, G. Kastenmüller, W. Römisch-Margl, M. Brugger, A. Peters, C. Meisinger, K. Strauch, C. Hengstenberg, P. Pagel, F. Huber, R. P. Mohney, H. Grallert, T. Illig, J. Adamski, M. Waldenberger, C. Gieger, and K. Suhre, "Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits," Hum. Mol. Genet., vol. 23, pp. 534–545, Jan. 2014.

[29] K. Suhre, S.-Y. Shin, A.-K. Petersen, R. P. Mohney, D. Meredith, B. Wägele, E. Altmaier, CARDIoGRAM, P. Deloukas, J. Erdmann, E. Grundberg, C. J. Hammond, M. H. de Angelis, G. Kastenmüller, A. Köttgen, F. Kronenberg, M. Mangino, C. Meisinger, T. Meitinger, H.-W. Mewes, M. V. Milburn, C. Prehn, J. Raffler, J. S. Ried, W. Römisch-Margl, N. J. Samani, K. S. Small, H.-E. Wichmann, G. Zhai, T. Illig, T. D. Spector, J. Adamski, N. Soranzo, and C. Gieger, "Human metabolic individuality in biomedical and pharmaceutical research," Nature, vol. 477, pp. 54–60, Aug. 2011.

[30] K. Suhre and C. Gieger, "Genetic variation in metabolic phenotypes: study designs and applications," Nat. Rev. Genet., vol. 13, pp. 759–769, Nov. 2012.

[31] A.-K. Petersen, J. Krumsiek, B. Wägele, F. J. Theis, H.-E. Wichmann, C. Gieger, and K. Suhre, "On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies," BMC Bioinformatics, vol. 13, p. 120, June 2012.

[32] M. Jaremek, Z. Yu, M. Mangino, K. Mittelstrass, C. Prehn, P. Singmann, T. Xu, N. Dahmen, K. M. Weinberger, K. Suhre, A. Peters, A. Döring, H. Hauner, J. Adamski, T. Illig, T. D. Spector, and R. Wang-Sattler, "Alcohol-induced metabolomic differences in humans," Transl. Psychiatry, vol. 3, p. e276, July 2013.

[33] O. Shaham, R. Wei, T. J. Wang, C. Ricciardi, G. D. Lewis, R. S. Vasan, S. A. Carr, R. Thadhani, R. E. Gerszten, and V. K. Mootha, "Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity," Mol. Syst. Biol., vol. 4, p. 214, Aug. 2008.

[34] R. J. DeBerardinis and C. B. Thompson, "Cellular metabolism and disease: what do metabolic outliers teach us?," Cell, vol. 148, pp. 1132–1144, Mar. 2012.

[35] M. M. Rinschen, J. Ivanisevic, M. Giera, and G. Siuzdak, "Identification of bioactive metabolites using activity metabolomics," Nat. Rev. Mol. Cell Biol., vol. 20, pp. 353–367, June 2019.

[36] J. S. Hawe, F. J. Theis, and M. Heinig, "Inferring interaction networks from Multi-Omics data," Front. Genet., vol. 10, p. 535, June 2019.

[37] J. Yan, S. L. Risacher, L. Shen, and A. J. Saykin, "Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data," Brief. Bioinform., vol. 19, pp. 1370–1381, Nov. 2018.

[38] C. Meng, O. A. Zeleznik, G. G. Thallinger, B. Kuster, A. M. Gholami, and A. C. Culhane, "Dimension reduction techniques for the integrative analysis of multi-omics data," Brief. Bioinform., vol. 17, pp. 628–641, July 2016.

[39] B. Mirza, W. Wang, J. Wang, H. Choi, N. C. Chung, and P. Ping, "Machine learning and integrative analysis of biomedical big data," Genes, vol. 10, Jan. 2019.

[40] C. Wu, F. Zhou, J. Ren, X. Li, Y. Jiang, and S. Ma, "A selective review of Multi-Level omics data integration using variable selection," High Throughput, vol. 8, p. 4, Jan. 2019.

[41] Y. Li, F.-X. Wu, and A. Ngom, "A review on machine learning principles for multi-view biological data integration," Brief. Bioinform., vol. 19, pp. 325–340, Mar. 2018.

[42] S. H. Chu, M. Huang, R. S. Kelly, E. Benedetti, J. K. Siddiqui, O. A. Zeleznik, A. Pereira, D. Herrington, C. E. Wheelock, J. Krumsiek, M. McGeachie, S. C. Moore, P. Kraft, E. Mathé, J. Lasky-Su, and Consortium of Metabolomics Studies Statistics Working Group, "Integration of metabolomic and other omics data in Population-Based study designs: An epidemiological perspective," Metabolites, vol. 9, June 2019.

[43] T. Eicher, G. Kinnebrew, A. Patt, K. Spencer, K. Ying, Q. Ma, R. Machiraju, and E. A. Mathé, "Metabolomics and Multi-Omics integration: A survey of computational methods and resources," Metabolites, vol. 10, p. 202, May 2020.

[44] M. S. Ghaemi, D. B. DiGiulio, K. Contrepois, B. Callahan, T. T. M. Ngo, B. Lee-McMullen, B. Lehallier, A. Robaczewska, D. Mcilwain, Y. Rosenberg-Hasson, R. J. Wong, C. Quaintance, A. Culos, N. Stanley, A. Tanada, A. Tsai, D. Gaudilliere, E. Ganio, X. Han, K. Ando, L. McNeil, M. Tingle, P. Wise, I. Maric, M. Sirota, T. Wyss-Coray, V. D. Winn, M. L. Druzin, R. Gibbs, G. L. Darmstadt, D. B. Lewis, V. Partovi Nia, B. Agard, R. Tibshirani, G. Nolan, M. P. Snyder, D. A. Relman, S. R. Quake, G. M. Shaw, D. K. Stevenson, M. S. Angst, B. Gaudilliere, and N. Aghaeepour, "Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy," Bioinformatics, vol. 35, pp. 95–103, Jan. 2019.

[45] L. Xicota, F. Ichou, F.-X. Lejeune, B. Colsch, A. Tenenhaus, I. Leroy, G. Fontaine, M. Lhomme, H. Bertin, M.-O. Habert, S. Epelbaum, B. Dubois, F. Mochel, M.-C. Potier, and INSIGHT study group, "Multi-omics signature of brain amyloid deposition in asymptomatic individuals at-risk for alzheimer's disease: The INSIGHT-preAD study," EBioMedicine, vol. 47, pp. 518–528, Sept. 2019.

[46] E. Borgan, B. Sitter, O. C. Lingjærde, H. Johnsen, S. Lundgren, T. F. Bathen, T. Sørlie, A.-L. Børresen-Dale, and I. S. Gribbestad, "Merging transcriptomics and metabolomics–advances in breast cancer profiling," BMC Cancer, vol. 10, p. 628, Nov. 2010.

[47] A. Singh, C. P. Shannon, B. Gautier, F. Rohart, M. Vacher, S. J. Tebbutt, and K.-A. Lê Cao, "DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays," Bioinformatics, vol. 35, pp. 3055–3062, Sept. 2019.

[48] C. Qiu, F. Yu, K. Su, Q. Zhao, L. Zhang, C. Xu, W. Hu, Z. Wang, L. Zhao, Q. Tian, Y. Wang, H. Deng, and H. Shen, "Multi-omics data integration for identifying osteoporosis biomarkers and their biological interaction and causal mechanisms," iScience, vol. 23, p. 100847, Feb. 2020.

[49] J. Zierer, T. Pallister, P.-C. Tsai, J. Krumsiek, J. T. Bell, G. Lauc, T. D. Spector, C. Menni, and G. Kastenmüller, "Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model," Sci. Rep., vol. 6, p. 37646, Nov. 2016.

[50] M. Altenbuchinger, H. U. Zacharias, S. Solbrig, A. Schäfer, M. Büyüközkan, U. T. Schultheiß, F. Kotsis, A. Köttgen, R. Spang, P. J. Oefner, J. Krumsiek, and W. Gronwald, "A multi-source data integration approach reveals novel associations between metabolites and renal outcomes in the german chronic kidney disease study," Sci. Rep., vol. 9, p. 13954, Sept. 2019.

[51] Q. Yao, Y. Xu, H. Yang, D. Shang, C. Zhang, Y. Zhang, Z. Sun, X. Shi, L. Feng, J. Han, F. Su, C. Li, and X. Li, "Global prioritization of disease candidate metabolites based on a multi-omics composite network," Sci. Rep., vol. 5, p. 17201, Nov. 2015.

[52] M. Bersanelli, E. Mosca, D. Remondini, E. Giampieri, C. Sala, G. Castellani, and L. Milanesi, "Methods for the integration of multi-omics data: mathematical aspects," BMC Bioinformatics, vol. 17 Suppl 2, p. 15, Jan. 2016.

[53] D. J. Beale, K. A. Kouremenos, and E. A. Palombo, Microbial Metabolomics, vol. Chapter 10. Springer International Publishing, 2016.

[54] V. Voillet, P. Besse, L. Liaubet, M. San Cristobal, and I. González, "Handling missing rows in multi-omics data integration: multiple imputation in multiple factor analysis framework," BMC Bioinformatics, vol. 17, p. 402, Oct. 2016.

[55] M. W. Weiner and Adni, "The ADNI initiative: review of paper published since its inception," Alzheimers. Dement., vol. 9, no. 5, pp. e111–e194, 2013.

[56] S. Nakagawa and R. P. Freckleton, "Missing inaction: the dangers of ignoring missing data," Trends Ecol. Evol., vol. 23, pp. 592–596, Nov. 2008.

[57] R. A. van den Berg, H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf, "Centering, scaling, and transformations: improving the biological information content of metabolomics data," BMC Genomics, vol. 7, p. 142, June 2006.

[58] R. Cavill, D. Jennen, J. Kleinjans, and J. J. Briedé, "Transcriptomic and metabolomic data integration," Brief. Bioinform., vol. 17, pp. 891–901, Sept. 2016.

[59] R. E. Bellman, "Adaptive control processes," Princeton University Press, 1961.

[60] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?," in Database Theory — ICDT'99, pp. 217–235, Springer Berlin Heidelberg, 1999.

[61] P. L., H. E., and L. H., "Subspace clustering for high dimensional data: A review," SIGKDD Explorations, Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining, vol. 6, no. 1, p. 90, 2004.

[62] K. T. Do, D. J. N.-P. Rasp, G. Kastenmüller, K. Suhre, and J. Krumsiek, "MoDentify: phenotype-driven module identification in metabolomics networks at different resolutions," Bioinformatics, vol. 35, pp. 532–534, Feb. 2019.

[63] J. Krumsiek, K. Mittelstrass, K. T. Do, F. Stückler, J. Ried, J. Adamski, A. Peters, T. Illig, F. Kronenberg, N. Friedrich, M. Nauck, M. Pietzner, D. O. Mook-Kanamori, K. Suhre, C. Gieger, H. Grallert, F. J. Theis, and G. Kastenmüller, "Gender-specific pathway differences in the human serum metabolome," Metabolomics, vol. 11, pp. 1815–1833, Aug. 2015.

[64] K. T. Do, M. Pietzner, D. J. Rasp, N. Friedrich, M. Nauck, T. Kocher, K. Suhre, D. O. Mook-Kanamori, G. Kastenmüller, and J. Krumsiek, "Phenotype-driven identification of modules in a hierarchical map of multifluid metabolic correlations," NPJ Syst Biol Appl, vol. 3, p. 28, Sept. 2017.

[65] J. Krumsiek, J. Bartel, and F. J. Theis, "Computational approaches for systems metabolomics," Curr. Opin. Biotechnol., vol. 39, pp. 198–206, June 2016.

[66] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," Chemometrics Intellig. Lab. Syst., vol. 2, pp. 37–52, Aug. 1987.

[67] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means clustering algorithm," J. R. Stat. Soc. Ser. C Appl. Stat., vol. 28, no. 1, pp. 100–108, 1979.

[68] S. C. Johnson, "Hierarchical clustering schemes," Psychometrika, vol. 32, pp. 241–254, Sept. 1967.

[69] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res., vol. 3, no. Mar, pp. 1157–1182, 2003.

[70] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," BMC Bioinformatics, vol. 9, p. 559, Dec. 2008.

[71] S. Wahl, S. Vogt, F. Stückler, J. Krumsiek, J. Bartel, T. Kacprowski, K. Schramm, M. Carstensen, W. Rathmann, M. Roden, C. Jourdan, A. J. Kangas, P. Soininen, M. Ala-Korpela, U. Nöthlings, H. Boeing, F. J. Theis, C. Meisinger, M. Waldenberger, K. Suhre, G. Homuth, C. Gieger, G. Kastenmüller, T. Illig, J. Linseisen, A. Peters, H. Prokisch, C. Herder, B. Thorand, and H. Grallert, "Multi-omic signature of body weight change: results from a population-based cohort study," BMC Med., vol. 13, p. 48, Mar. 2015.

[72] R. L. Costa, M. Boroni, and M. A. Soares, "Distinct co-expression networks using multi-omic data reveal novel interventional targets in HPV-positive and negative head-and-neck squamous cell cancer," Sci. Rep., vol. 8, p. 15254, Oct. 2018.

[73] H. K. Pedersen, S. K. Forslund, V. Gudmundsdottir, A. Ø. Petersen, F. Hildebrand, T. Hyötyläinen, T. Nielsen, T. Hansen, P. Bork, S. D. Ehrlich, S. Brunak, M. Oresic, O. Pedersen, and H. B. Nielsen, "A computational framework to integrate high-throughput '-omics' datasets for the identification of potential mechanistic links," Nat. Protoc., vol. 13, pp. 2781–2800, Dec. 2018.

[74] R. Hernández-de Diego, S. Tarazona, C. Martínez-Mira, L. Balzano-Nogueira, P. Furió-Tarí, G. J. Pappas, Jr, and A. Conesa, "PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data," Nucleic Acids Res., vol. 46, pp. W503–W509, July 2018.

[75] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson, and M. J. Herrgard, "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox," Nat. Protoc., vol. 2, no. 3, pp. 727–738, 2007.

[76] Cancer Genome Atlas Research Network, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. M. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, and J. M. Stuart, "The cancer genome atlas Pan-Cancer analysis project," Nat. Genet., vol. 45, pp. 1113–1120, Oct. 2013.

[77] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, and C. von Mering, "STRING v10: protein-protein interaction networks, integrated over the tree of life," Nucleic Acids Res., vol. 43, pp. D447–52, Jan. 2015.

[78] B. Boeckmann, A. Bairoch, R. Apweiler, M.-C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, and M. Schneider, "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003," Nucleic Acids Res., vol. 31, pp. 365–370, Jan. 2003.

[79] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. v. Mering, "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," Nucleic Acids Res., vol. 47, pp. D607–D613, Jan. 2019.

[80] M. Sud, E. Fahy, D. Cotter, A. Brown, E. A. Dennis, C. K. Glass, A. H. Merrill, Jr, R. C. Murphy, C. R. H. Raetz, D. W. Russell, and S. Subramaniam, "LMSD: LIPID MAPS structure database," Nucleic Acids Res., vol. 35, pp. D527–32, Jan. 2007.

[81] M. Kanehisa, "Toward understanding the origin and evolution of cellular organisms," Protein Sci., vol. 28, pp. 1947–1951, Nov. 2019.

[82] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," Nucleic Acids Res., vol. 28, pp. 27–30, Jan. 2000.

[83] M. Kanehisa, Y. Sato, M. Furumichi, K. Morishima, and M. Tanabe, "New approach for understanding genome variations in KEGG," Nucleic Acids Res., vol. 47, pp. D590–D595, Jan. 2019.

[84] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein, "Reactome: a knowledgebase of biological pathways," Nucleic Acids Res., vol. 33, pp. D428–32, Jan. 2005.

[85] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C. D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio, "The reactome pathway knowledgebase," Nucleic Acids Res., vol. 46, pp. D649–D655, Jan. 2018.

[86] I. Thiele, N. Swainston, R. M. T. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottir, M. L. Mo, O. Rolfsson, M. D. Stobbe, S. G. Thorleifsson, R. Agren, C. Bölling, S. Bordel, A. K. Chavali, P. Dobson, W. B. Dunn, L. Endler, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J. J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. Le Novère, N. Malys, A. Mazein, J. A. Papin, N. D. Price, E. Selkov, Sr, M. I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, J. H. G. M. van Beek, D. Weichart, I. Goryanin, J. Nielsen, H. V. Westerhoff, D. B. Kell, P. Mendes, and B. Ø. Palsson, "A community-driven global reconstruction of human metabolism," Nat. Biotechnol., vol. 31, pp. 419–425, May 2013.

[87] N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Ø. Palsson, "Global reconstruction of the human metabolic network based on genomic and bibliomic data," Proc. Natl. Acad. Sci. U. S. A., vol. 104, pp. 1777–1782, Feb. 2007.

[88] E. Brunk, S. Sahoo, D. C. Zielinski, A. Altunkaya, A. Dräger, N. Mih, F. Gatto, A. Nilsson, G. A. Preciat Gonzalez, M. K. Aurich, A. Prlić, A. Sastry, A. D. Danielsdottir, A. Heinken, A. Noronha, P. W. Rose, S. K. Burley, R. M. T. Fleming, J. Nielsen, I. Thiele, and B. O. Palsson, "Recon3D enables a three-dimensional view of gene variation in human metabolism," Nat. Biotechnol., vol. 36, pp. 272–281, Mar. 2018.

[89] S. Durinck, P. T. Spellman, E. Birney, and W. Huber, "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomart," Nat. Protoc., vol. 4, pp. 1184–1191, July 2009.

[90] J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D. S. Wishart, and J. Xia, "MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis," Nucleic Acids Res., vol. 46, pp. W486–W494, July 2018.

[91] J. Chong and J. Xia, "MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data," Bioinformatics, vol. 34, pp. 4313–4314, Dec. 2018.

[92] N. Pham, R. G. A. van Heck, J. C. J. van Dam, P. J. Schaap, E. Saccenti, and M. Suarez-Diez, "Consistency, inconsistency, and ambiguity of metabolite names in biochemical databases used for Genome-Scale metabolic modelling," Metabolites, vol. 9, Feb. 2019.

[93] J. D. Quell, W. Römisch-Margl, M. Haid, J. Krumsiek, T. Skurk, A. Halama, N. Stephan, J. Adamski, H. Hauner, D. Mook-Kanamori, R. P. Mohney, H. Daniel, K. Suhre, and G. Kastenmüller, "Characterization of bulk phosphatidylcholine compositions in human plasma using Side-Chain resolving lipidomics," Metabolites, vol. 9, June 2019.

[94] J. P. Koelmel, C. Z. Ulmer, C. M. Jones, R. A. Yost, and J. A. Bowden, "Common cases of improper lipid annotation using high-resolution tandem mass spectrometry data and corresponding limitations in biological interpretation," Biochim. Biophys. Acta Mol. Cell Biol. Lipids, vol. 1862, pp. 766–770, Aug. 2017.

[95] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, and P. D. Karp, "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases," Nucleic Acids Res., vol. 44, pp. D471–80, Jan. 2016.

[96] I. Rodchenkov, O. Babur, A. Luna, B. A. Aksoy, J. V. Wong, D. Fong, M. Franz, M. C. Siper, M. Cheung, M. Wrana, H. Mistry, L. Mosier, J. Dlin, Q. Wen, C. O'Callaghan, W. Li, G. Elder, P. T. Smith, C. Dallago, E. Cerami, B. Gross, U. Dogrusoz, E. Demir, G. D. Bader, and C. Sander, "Pathway commons 2019 update: integration, analysis and exploration of pathway data," Nucleic Acids Res., vol. 48, pp. D489–D497, Jan. 2020.

[97] D. N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, N. Nunes, J. Mélius, E. Cirillo, S. L. Coort, D. Digles, F. Ehrhart, P. Giesbertz, M. Kalafati, M. Martens, R. Miller, K. Nishida, L. Rieswijk, A. Waagmeester, L. M. T. Eijssen, C. T. Evelo, A. R. Pico, and E. L. Willighagen, "WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research," Nucleic Acids Res., vol. 46, pp. D661–D667, Jan. 2018.

[98] T. Kelder, A. R. Pico, K. Hanspers, M. P. van Iersel, C. Evelo, and B. R. Conklin, "Mining biological pathways using WikiPathways web services," PLoS One, vol. 4, p. e6447, July 2009.

[99] R. T. Pillich, J. Chen, V. Rynkov, D. Welker, and D. Pratt, "NDEx: A community resource for sharing and publishing of biological networks," in Protein Bioinformatics, Methods in molecular biology (Clifton, N.J.), pp. 271–301, New York, NY: Springer New York, 2017.

[100] D. Pratt, J. Chen, R. Pillich, V. Rynkov, A. Gary, B. Demchak, and T. Ideker, "NDEx 2.0: A clearinghouse for research on cancer pathways," Cancer Res., vol. 77, pp. e58–e61, Nov. 2017.

[101] D. Pratt, J. Chen, D. Welker, R. Rivas, R. Pillich, V. Rynkov, K. Ono, C. Miello, L. Hicks, S. Szalma, A. Stojmirovic, R. Dobrin, M. Braxenthaler, J. Kuentzer, B. Demchak, and T. Ideker, "NDEx, the network data exchange," Cell Syst, vol. 1, pp. 302–305, Oct. 2015.

[102] G. Zhou and J. Xia, "Using OmicsNet for network integration and 3D visualization," Curr. Protoc. Bioinformatics, vol. 65, p. e69, Mar. 2019.

[103] L. Cottret, C. Frainay, M. Chazalviel, F. Cabanettes, Y. Gloaguen, E. Camenen, B. Merlet, S. Heux, J.-C. Portais, N. Poupin, F. Vinson, and F. Jourdan, "MetExplore: collaborative edition and exploration of metabolic networks," Nucleic Acids Res., vol. 46, pp. W495–W502, July 2018.

[104] A. Kamburov, C. Wierling, H. Lehrach, and R. Herwig, "ConsensusPathDB–a database for integrating human functional interaction networks," Nucleic Acids Res., vol. 37, pp. D623–8, Jan. 2009.

[105] D. Domingo-Fernández, S. Mubeen, J. Marín-Llaó, C. T. Hoyt, and M. Hofmann-Apitius, "PathMe: merging and exploring mechanistic pathway knowledge," BMC Bioinformatics, vol. 20, p. 243, May 2019.

[106] A. Karnovsky, T. Weymouth, T. Hull, V. G. Tarcea, G. Scardoni, C. Laudanna, M. A. Sartor, K. A. Stringer, H. V. Jagadish, C. Burant, B. Athey, and G. S. Omenn, "Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data," Bioinformatics, vol. 28, pp. 373–380, Feb. 2012.

[107] S. Basu, W. Duren, C. R. Evans, C. F. Burant, G. Michailidis, and A. Karnovsky, "Sparse network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data," Bioinformatics, vol. 33, pp. 1545–1553, May 2017.

[108] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," Genome Res., vol. 13, pp. 2498–2504, Nov. 2003.

[109] J. Xia and D. S. Wishart, "MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data," Nucleic Acids Res., vol. 38, pp. W71–7, July 2010.

[110] R. A. Fisher, "Statistical methods for research workers," in Breakthroughs in Statistics: Methodology and Distribution (S. Kotz and N. L. Johnson, eds.), pp. 66–70, New York, NY: Springer New York, 1992.

[111] S. A. Stouffer, E. A. Suchman, L. C. Devinney, S. A. Star, and R. M. Williams, Jr, "The american soldier: Adjustment during army life. (studies in social psychology in world war II), vol. 1," vol. 1, p. 599, 1949.

[112] T. Lipták, "On the combination of independent tests," Magyar Tud Akad Mat Kutato Int Kozl, vol. 3, pp. 171–197, 1958.

[113] A. Kamburov, R. Cavill, T. M. D. Ebbels, R. Herwig, and H. C. Keun, "Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA," Bioinformatics, vol. 27, pp. 2917–2918, Oct. 2011.

[114] J. Xia, I. V. Sinelnikov, B. Han, and D. S. Wishart, "MetaboAnalyst 3.0–making metabolomics more meaningful," Nucleic Acids Res., vol. 43, pp. W251–7, July 2015.

[115] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," Proc. Natl. Acad. Sci. U. S. A., vol. 102, pp. 15545–15550, Oct. 2005.

[116] M. R. Molenaar, A. Jeucken, T. A. Wassenaar, C. H. A. van de Lest, J. F. Brouwers, and J. B. Helms, "LION/web: a web-based ontology enrichment tool for lipidomic data analysis," Gigascience, vol. 8, pp. 1–10, June 2019.

[117] M. List, N. Alcaraz, M. Dissing-Hansen, H. J. Ditzel, J. Mollenhauer, and J. Baumbach, "KeyPathwayMinerWeb: online multi-omics network enrichment," Nucleic Acids Res., vol. 44, pp. W98–W104, July 2016.

[118] N. Alcaraz, J. Pauling, R. Batra, E. Barbosa, A. Junge, A. G. L. Christensen, V. Azevedo, H. J. Ditzel, and J. Baumbach, "KeyPathwayMiner 4.0: condition-specific pathway analysis by combining multiple omics studies and networks with cytoscape," BMC Syst. Biol., vol. 8, p. 99, Aug. 2014.

[119] R. Batra, N. Alcaraz, K. Gitzhofer, J. Pauling, H. J. Ditzel, M. Hellmuth, J. Baumbach, and M. List, "On the performance of de novo pathway enrichment," NPJ Syst Biol Appl, vol. 3, p. 6, Mar. 2017.

[120] M. Soerensen, D. M. Hozakowska-Roszkowska, M. Nygaard, M. J. Larsen, V. Schwämmle, K. Christensen, L. Christiansen, and Q. Tan, "A Genome-Wide integrative association study of DNA methylation and gene expression data and later life cognitive functioning in monozygotic twins," Front. Neurosci., vol. 14, p. 233, Apr. 2020.

[121] E. Stalidzans, M. Zanin, P. Tieri, F. Castiglione, A. Polster, S. Scheiner, J. Pahle, B. Stres, M. List, J. Baumbach, M. Lautizi, K. Van Steen, and H. H. H. W. Schmidt, "Mechanistic modeling and multiscale applications for precision medicine: Theory and practice," Network and Systems Medicine, vol. 3, pp. 36–56, Aug. 2020.

[122] J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?," Nat. Biotechnol., vol. 28, pp. 245–248, Mar. 2010.

[123] I. Thiele and B. Ø. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction," Nat. Protoc., vol. 5, pp. 93–121, Jan. 2010.

[124] L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S. M. Keating, V. Vlasov, S. Magnusdóttir, C. Y. Ng, G. Preciat, A. Žagare, S. H. J. Chan, M. K. Aurich, C. M. Clancy, J. Modamio, J. T. Sauls, A. Noronha, A. Bordbar, B. Cousins, D. C. El Assal, L. V. Valcarcel, I. Apaolaza, S. Ghaderi, M. Ahookhosh, M. Ben Guebila, A. Kostromins, N. Sompairac, H. M. Le, D. Ma, Y. Sun, L. Wang, J. T. Yurkovich, M. A. P. Oliveira, P. T. Vuong, L. P. El Assal, I. Kuperstein, A. Zinovyev, H. S. Hinton, W. A. Bryant, F. J. Aragón Artacho, F. J. Planes, E. Stalidzans, A. Maass, S. Vempala, M. Hucka, M. A. Saunders, C. D. Maranas, N. E. Lewis, T. Sauter, B. Ø. Palsson, I. Thiele, and R. M. T. Fleming, "Creation and analysis of biochemical constraint-based models using the COBRA toolbox v.3.0," Nat. Protoc., vol. 14, pp. 639–702, Mar. 2019.

[125] A. Bordbar, J. M. Monk, Z. A. King, and B. O. Palsson, "Constraint-based models predict metabolic and associated cellular functions," Nat. Rev. Genet., vol. 15, pp. 107–120, Feb. 2014.

[126] K. R. Patil and J. Nielsen, "Uncovering transcriptional regulation of metabolism by using metabolic network topology," Proc. Natl. Acad. Sci. U. S. A., vol. 102, pp. 2685–2689, Feb. 2005.

[127] J. S. Cho, C. Gu, T. H. Han, J. Y. Ryu, and S. Y. Lee, "Reconstruction of context-specific genome-scale metabolic models using multiomics data to study metabolic rewiring," Current Opinion in Systems Biology, vol. 15, pp. 1–11, June 2019.

[128] J. L. Reed, "Shrinking the metabolic solution space using experimental datasets," PLoS Comput. Biol., vol. 8, p. e1002662, Aug. 2012.

[129] N. Vlassis, M. P. Pacheco, and T. Sauter, "Fast reconstruction of compact context-specific metabolic network models," PLoS Comput. Biol., vol. 10, p. e1003424, Jan. 2014.

[130] A. Bordbar, J. T. Yurkovich, G. Paglia, O. Rolfsson, Ó. E. Sigurjónsson, and B. O. Palsson, "Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics," Sci. Rep., vol. 7, p. 46249, Apr. 2017.

[131] H. Zur, E. Ruppin, and T. Shlomi, "iMAT: an integrative metabolic analysis tool," Bioinformatics, vol. 26, pp. 3140–3142, Dec. 2010.

[132] S. A. Becker and B. O. Palsson, "Context-specific metabolic networks are consistent with experiments," PLoS Comput. Biol., vol. 4, p. e1000082, May 2008.

[133] R. Saha, A. Chowdhury, and C. D. Maranas, "Recent advances in the reconstruction of metabolic models and integration of omics data," Curr. Opin. Biotechnol., vol. 29, pp. 39–45, Oct. 2014.

[134] R. Agren, A. Mardinoglu, A. Asplund, C. Kampf, M. Uhlen, and J. Nielsen, "Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling," Mol. Syst. Biol., vol. 10, p. 721, Mar. 2014.

[135] I. Thiele, S. Sahoo, A. Heinken, L. Heirendt, M. K. Aurich, A. Noronha, and R. M. T. Fleming, "When metabolism meets physiology: Harvey and harvetta," bioRxiv, Jan. 2018.

[136] H.-B. Shen and K.-C. Chou, "Ensemble classifier for protein fold pattern recognition," Bioinformatics, vol. 22, pp. 1717–1722, July 2006.

[137] L. Rokach, "Ensemble-based classifiers," Artificial Intelligence Review, vol. 33, pp. 1–39, Feb. 2010.

[138] C. Fawcett and H. H. Hoos, "Analysing differences between algorithm configurations through ablation," Journal of Heuristics, vol. 22, pp. 431–458, Aug. 2016.

[139] C. M. Miles and M. Wayne, "Quantitative trait locus (QTL) analysis," Nature Education 1 (1), vol. 208, 2008.

[140] J. N. Hirschhorn and M. J. Daly, "Genome-wide association studies for common diseases and complex traits," Nat. Rev. Genet., vol. 6, pp. 95–108, Feb. 2005.

[141] M. I. McCarthy, G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. A. Ioannidis, and J. N. Hirschhorn, "Genome-wide association studies for complex traits: consensus, uncertainty and challenges," Nat. Rev. Genet., vol. 9, pp. 356–369, May 2008.

[142] V. Tam, N. Patel, M. Turcotte, Y. Bossé, G. Paré, and D. Meyre, "Benefits and limitations of genome-wide association studies," Nat. Rev. Genet., vol. 20, pp. 467–484, Aug. 2019.

[143] GTEx Consortium, "The GTEx consortium atlas of genetic regulatory effects across human tissues," Science, vol. 369, pp. 1318–1330, Sept. 2020.

[144] GTEx Consortium, "The Genotype-Tissue expression (GTEx) project," Nat. Genet., vol. 45, pp. 580–585, June 2013.

[145] B. B. Sun, J. C. Maranville, J. E. Peters, D. Stacey, J. R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M. A. Kamat, B. P. Prins, S. K. Wilcox, E. S. Zimmerman, A. Chi, N. Bansal, S. L. Spain, A. M. Wood, N. W. Morrell, J. R. Bradley, N. Janjic, D. J. Roberts, W. H. Ouwehand, J. A. Todd, N. Soranzo, K. Suhre, D. S. Paul, C. S. Fox, R. M. Plenge, J. Danesh, H. Runz, and A. S. Butterworth, "Genomic atlas of the human plasma proteome," Nature, vol. 558, pp. 73–79, June 2018.

[146] K. Suhre, M. Arnold, A. M. Bhagwat, R. J. Cotton, R. Engelke, J. Raffler, H. Sarwath, G. Thareja, A. Wahl, R. K. DeLisle, L. Gold, M. Pezer, G. Lauc, M. A. El-Din Selim, D. O. Mook-Kanamori, E. K. Al-Dous, Y. A. Mohamoud, J. Malek, K. Strauch, H. Grallert, A. Peters, G. Kastenmüller, C. Gieger, and J. Graumann, "Connecting genetic risk to disease end points through the human blood plasma proteome," Nat. Commun., vol. 8, p. 14357, Feb. 2017.

[147] C. Sabatti, S. K. Service, A.-L. Hartikainen, A. Pouta, S. Ripatti, J. Brodsky, C. G. Jones, N. A. Zaitlen, T. Varilo, M. Kaakinen, U. Sovio, A. Ruokonen, J. Laitinen, E. Jakkula, L. Coin, C. Hoggart, A. Collins, H. Turunen, S. Gabriel, P. Elliot, M. I. McCarthy, M. J. Daly, M.-R. Järvelin, N. B. Freimer, and L. Peltonen, "Genome-wide association analysis of metabolic traits in a birth cohort from a founder population," Nat. Genet., vol. 41, pp. 35–46, Jan. 2009.

[148] M. Arnold, J. Raffler, A. Pfeufer, K. Suhre, and G. Kastenmüller, "SNiPA: an interactive, genetic variant-centered annotation browser," Bioinformatics, vol. 31, pp. 1334–1336, Apr. 2015.

[149] F. Hormozdiari, M. van de Bunt, A. V. Segrè, X. Li, J. W. J. Joo, M. Bilow, J. H. Sul, S. Sankararaman, B. Pasaniuc, and E. Eskin, "Colocalization of GWAS and eQTL signals detects target genes," Am. J. Hum. Genet., vol. 99, pp. 1245–1260, Dec. 2016.

[150] C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, and V. Plagnol, "Bayesian test for colocalisation between pairs of genetic association studies using summary statistics," PLoS Genet., vol. 10, p. e1004383, May 2014.

[151] M. J. Bonder, R. Luijk, D. V. Zhernakova, M. Moed, P. Deelen, M. Vermaat, M. van Iterson, F. van Dijk, M. van Galen, J. Bot, R. C. Slieker, P. M. Jhamai, M. Verbiest, H. E. D. Suchiman, M. Verkerk, R. van der Breggen, J. van Rooij, N. Lakenberg, W. Arindrarto, S. M. Kielbasa, I. Jonkers, P. van 't Hof, I. Nooren, M. Beekman, J. Deelen, D. van Heemst, A. Zhernakova, E. F. Tigchelaar, M. A. Swertz, A. Hofman, A. G. Uitterlinden, R. Pool, J. van Dongen, J. J. Hottenga, C. D. A. Stehouwer, C. J. H. van der Kallen, C. G. Schalkwijk, L. H. van den Berg, E. W. van Zwet, H. Mei, Y. Li, M. Lemire, T. J. Hudson, BIOS Consortium, P. E. Slagboom, C. Wijmenga, J. H. Veldink, M. M. J. van Greevenbroek, C. M. van Duijn, D. I. Boomsma, A. Isaacs, R. Jansen, J. B. J. van Meurs, P. A. C. 't Hoen, L. Franke, and B. T. Heijmans, "Disease variants alter transcription factor levels and methylation of their binding sites," Nat. Genet., vol. 49, pp. 131–138, Jan. 2017.

[152] L. Chen, B. Ge, F. P. Casale, L. Vasquez, T. Kwan, D. Garrido-Martín, S. Watt, Y. Yan, K. Kundu, S. Ecker, A. Datta, D. Richardson, F. Burden, D. Mead, A. L. Mann, J. M. Fernandez, S. Rowlston, S. P. Wilder, S. Farrow, X. Shao, J. J. Lambourne, A. Redensek, C. A. Albers, V. Amstislavskiy, S. Ashford, K. Berentsen, L. Bomba, G. Bourque, D. Bujold, S. Busche, M. Caron, S. H. Chen, W. Cheung, O. Delaneau, E. T. Dermitzakis, H. Elding, I. Colgiu, F. O. Bagger, P. Flicek, E. Habibi, V. Iotchkova, E. Janssen-Megens, B. Kim, H. Lehrach, E. Lowy, A. Mandoli, F. Matarese, M. T. Maurano, J. A. Morris, V. Pancaldi, F. Pourfarzad, K. Rehnstrom, A. Rendon, T. Risch, N. Sharifi, M. M. Simon, M. Sultan, A. Valencia, K. Walter, S. Y. Wang, M. Frontini, S. E. Antonarakis, L. Clarke, M. L. Yaspo, S. Beck, R. Guigo, D. Rico, J. H. A. Martens, W. H. Ouwehand, T. W. Kuijpers, D. S. Paul, H. G. Stunnenberg, O. Stegle, K. Downes, T. Pastinen, and N. Soranzo, Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells, vol. 167. 2016.

[153] C. L. Masters, R. Bateman, K. Blennow, C. C. Rowe, R. A. Sperling, and J. L. Cummings, "Alzheimer's disease," Nat Rev Dis Primers, vol. 1, p. 15056, Oct. 2015.

[154] E. Evangelou and J. P. A. Ioannidis, "Meta-analysis methods for genome-wide association studies and beyond," Nat. Rev. Genet., vol. 14, pp. 379–389, June 2013.

[155] D. B. Rubin, Multiple Imputation for Nonresponse in Surveys, vol. 81. John Wiley & Sons, June 2004.

[156] R. Argelaguet, B. Velten, D. Arnol, S. Dietrich, T. Zenz, J. C. Marioni, F. Buettner, W. Huber, and O. Stegle, "Multi-Omics factor analysis-a framework for unsupervised integration of multi-omics data sets," Mol. Syst. Biol., vol. 14, p. e8124, June 2018.

[157] R. Argelaguet, D. Arnol, D. Bredikhin, Y. Deloro, B. Velten, J. C. Marioni, and O. Stegle, "MOFA+: a probabilistic framework for comprehensive integration of structured single-cell data," bioRxiv, p. 837104, Nov. 2019.

[158] H. Abdi, L. J. Williams, D. Valentin, and M. Bennani-Dosse, "STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling," Wiley Interdiscip. Rev. Comput. Stat., vol. 4, pp. 124–167, Mar. 2012.

[159] L. Breiman, "Random forests," Mach. Learn., vol. 45, pp. 5–32, Oct. 2001.

[160] R. Tishbirani, "Regression shrinkage and selection via the lasso," 1996.

[161] D. V. Nguyen and D. M. Rocke, "Tumor classification by partial least squares using microarray gene expression data," Bioinformatics, vol. 18, pp. 39–50, Jan. 2002.

[162] A.-L. Boulesteix, "PLS dimension reduction for classification with microarray data," Stat. Appl. Genet. Mol. Biol., vol. 3, p. Article33, Nov. 2004.

[163] J. Schäfer and K. Strimmer, "An empirical bayes approach to inferring large-scale gene association networks," Bioinformatics, vol. 21, pp. 754–764, Mar. 2005.

[164] J. Krumsiek, K. Suhre, T. Illig, J. Adamski, and F. J. Theis, "Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data," BMC Syst. Biol., vol. 5, p. 21, Jan. 2011.

[165] M. Altenbuchinger, A. Weihs, J. Quackenbush, H. J. Grabe, and H. U. Zacharias, "Gaussian and mixed graphical models as (multi-)omics data analysis tools," Biochim. Biophys. Acta Gene Regul. Mech., vol. 1863, p. 194418, June 2020.

[166] S. L. Lauritzen, Graphical Models, vol. 17. Clarendon Press, May 1996.

[167] K. Mittelstrass, J. S. Ried, Z. Yu, J. Krumsiek, C. Gieger, C. Prehn, W. Roemisch-Margl, A. Polonikov, A. Peters, F. J. Theis, T. Meitinger, F. Kronenberg, S. Weidinger, H. E. Wichmann, K. Suhre, R. Wang-Sattler, J. Adamski, and T. Illig, "Discovery of sexual dimorphisms in metabolic and genetic biomarkers," PLoS Genet., vol. 7, p. e1002215, Aug. 2011.

[168] J. Krumsiek, K. Suhre, A. M. Evans, M. W. Mitchell, R. P. Mohney, M. V. Milburn, B. Wägele, W. Römisch-Margl, T. Illig, J. Adamski, C. Gieger, F. J. Theis, and G. Kastenmüller, "Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information," PLoS Genet., vol. 8, p. e1003005, Oct. 2012.

[169] S. Chen, D. M. Witten, and A. Shojaie, "Selection and estimation for mixed graphical models," Biometrika, vol. 102, pp. 47–64, Mar. 2015.

[170] J. Lee and T. Hastie, "Structure learning of mixed graphical models," in Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (C. M. Carvalho and P. Ravikumar, eds.), vol. 31 of Proceedings of Machine Learning Research, (Scottsdale, Arizona, USA), pp. 388–396, PMLR, 2013.

[171] B. Fellinghauer, P. Bühlmann, M. Ryffel, M. von Rhein, and J. D. Reinhardt, "Stable graphical model estimation with random forests for discrete, continuous, and mixed variables," Comput. Stat. Data Anal., vol. 64, pp. 132–152, Aug. 2013.

[172] F. Rohart, B. Gautier, A. Singh, and K.-A. Lê Cao, "mixomics: An R package for 'omics feature selection and multiple data integration," PLoS Comput. Biol., vol. 13, p. e1005752, Nov. 2017.

[173] J. S. Spicker, S. Brunak, K. S. Frederiksen, and H. Toft, "Integration of clinical chemistry, expression, and metabolite data leads to better toxicological class separation," Toxicol. Sci., vol. 102, pp. 444–454, Apr. 2008.

[174] B. Escofier and J. Pagès, "Multiple factor analysis (AFMULT package)," Comput. Stat. Data Anal., vol. 18, pp. 121–140, Aug. 1994.

[175] H. Abdi, L. J. Williams, and D. Valentin, "Multiple factor analysis: principal component analysis for multitable and multiblock data sets," Wiley Interdiscip. Rev. Comput. Stat., vol. 5, pp. 149–179, Mar. 2013.

[176] S. Sun, "A survey of multi-view machine learning," Neural Comput. Appl., vol. 23, pp. 2031–2038, Dec. 2013.

[177] J. Trygg, "O2-PLS for qualitative and quantitative analysis in multivariate calibration," J. Chemom., vol. 16, pp. 283–293, June 2002.

[178] J. Trygg and S. Wold, "O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter," J. Chemom., vol. 17, pp. 53–64, Jan. 2003.

[179] V. Srivastava, O. Obudulu, J. Bygdell, T. Löfstedt, P. Rydén, R. Nilsson, M. Ahnlund, A. Johansson, P. Jonsson, E. Freyhult, J. Qvarnström, J. Karlsson, M. Melzer, T. Moritz, J. Trygg, T. R. Hvidsten, and G. Wingsle, "OnPLS integration of transcriptomic, proteomic and metabolomic data shows multi-level oxidative stress responses in the cambium of transgenic hipi- superoxide dismutase populus plants," BMC Genomics, vol. 14, p. 893, Dec. 2013.

[180] T. Löfstedt and J. Trygg, "OnPLS-a novel multiblock method for the modelling of predictive and orthogonal variation," J. Chemom., vol. 25, no. 8, pp. 441–455, 2011.

[181] T. Löfstedt, D. Hoffman, and J. Trygg, "Global, local and unique decompositions in OnPLS for multiblock data analysis," Anal. Chim. Acta, vol. 791, pp. 13–24, Aug. 2013.

[182] S. N. Reinke, B. Galindo-Prieto, T. Skotare, D. I. Broadhurst, A. Singhania, D. Horowitz, R. Djukanović, T. S. C. Hinks, P. Geladi, J. Trygg, and C. E. Wheelock, "OnPLS-Based Multi-Block data integration: A multivariate approach to interrogating biological interactions in asthma," Anal. Chem., vol. 90, pp. 13400–13408, Nov. 2018.

[183] A. Tenenhaus and M. Tenenhaus, "Regularized generalized canonical correlation analysis," Psychometrika, vol. 76, pp. 257–284, Apr. 2011.

[184] M. C. C. Langenberg, M.-A. Hoogerwerf, J. P. R. Koopman, J. J. Janse, J. Kos-van Oosterhoud, C. Feijt, S. P. Jochems, C. J. de Dood, R. van Schuijlenburg, A. Ozir-Fazalalikhan, M. D. Manurung, E. Sartono, M. T. van der Beek, B. M. F. Winkel, P. H. Verbeek-Menken, K. A. Stam, F. W. B. van Leeuwen, P. Meij, A. van Diepen, L. van Lieshout, G. J. van Dam, P. L. A. M. Corstjens, C. H. Hokke, M. Yazdanbakhsh, L. G. Visser, and M. Roestenberg, "A controlled human schistosoma mansoni infection model to advance novel drugs, vaccines and diagnostics," Nat. Med., vol. 26, pp. 326–332, Mar. 2020.

[185] G. Cano-Sancho, M.-C. Alexandre-Gouabau, T. Moyon, A.-L. Royer, Y. Guitton, H. Billard, D. Darmaun, J.-C. Rozé, C.-Y. Boquien, B. Le Bizec, and J.-P. Antignac, "Simultaneous exploration of nutrients and pollutants in human milk and their impact on preterm infant growth: An integrative cross-platform approach," Environ. Res., vol. 182, p. 109018, Mar. 2020.

[186] C. T. Pekmez, M. W. Larsson, M. V. Lind, N. Vazquez Manjarrez, C. Yonemitsu, A. Larnkjaer, L. Bode, C. Mølgaard, K. F. Michaelsen, and L. O. Dragsted, "Breastmilk lipids and oligosaccharides influence branched Short-Chain fatty acid concentrations in infants with excessive weight gain," Mol. Nutr. Food Res., vol. 64, p. e1900977, Feb. 2020.

[187] M. Vidal, M. E. Cusick, and A.-L. Barabási, "Interactome networks and human disease," Cell, vol. 144, pp. 986–998, Mar. 2011.

[188] E. Sügis, J. Dauvillier, A. Leontjeva, P. Adler, V. Hindie, T. Moncion, V. Collura, R. Daudin, Y. Loe-Mie, Y. Herault, J.-C. Lambert, H. Hermjakob, T. Pupko, J.-C. Rain, I. Xenarios, J. Vilo, M. Simonneau, and H. Peterson, "HENA, heterogeneous network-based data set for alzheimer's disease," Sci Data, vol. 6, p. 151, Aug. 2019.

[189] R. Herwig, C. Hardt, M. Lienhard, and A. Kamburov, "Analyzing and interpreting genome data at the network level with ConsensusPathDB," Nat. Protoc., vol. 11, pp. 1889–1907, Oct. 2016.

[190] G. Zhou and J. Xia, "OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space," Nucleic Acids Res., vol. 46, pp. W514–W522, July 2018.

[191] R. Haas, A. Zelezniak, J. Iacovacci, S. Kamrad, S. Townsend, and M. Ralser, "Designing and interpreting 'multi-omic' experiments that may change our understanding of biology," Curr Opin Syst Biol, vol. 6, pp. 37–45, Dec. 2017.

[192] D. Merico, D. Gfeller, and G. D. Bader, "How to visually interpret biological data using networks," Nat. Biotechnol., vol. 27, pp. 921–924, Oct. 2009.

[193] V. Yoghourdjian, D. Archambault, S. Diehl, T. Dwyer, K. Klein, H. C. Purchase, and H.-Y. Wu, "Exploring the limits of complexity: A survey of empirical studies on graph visualisation," arXiv [cs.HC], vol. 2, pp. 264–282, Sept. 2018.

[194] V. Yoghourdjian, T. Dwyer, K. Klein, K. Marriott, and M. Wybrow, "Graph thumbnails: Identifying and comparing multiple graphs at a glance," IEEE Trans. Vis. Comput. Graph., vol. 24, pp. 3081–3095, Dec. 2018.

[195] M. Krzywinski, I. Birol, S. J. M. Jones, and M. A. Marra, "Hive plots–rational approach to visualizing networks," Brief. Bioinform., vol. 13, pp. 627–644, Sept. 2012.

[196] F. McGee, M. Ghoniem, G. Melançon, B. Otjacques, and B. Pinaud, "The state of the art in multilayer network visualization," Comput. Graph. Forum, vol. 38, pp. 125–149, Sept. 2019.

[197] C. Meng, B. Kuster, A. C. Culhane, and A. M. Gholami, "A multivariate approach to the integration of multi-omics datasets," BMC Bioinformatics, vol. 15, p. 162, May 2014.

[198] X. Ge, V. K. Raghu, P. K. Chrysanthis, and P. V. Benos, "CausalMGM: an interactive web-based causal discovery tool," Nucleic Acids Res., vol. 48, pp. W597–W602, July 2020.

[199] K. Uppal, C. Ma, Y.-M. Go, D. P. Jones, and J. Wren, "xMWAS: a data-driven integration and differential network analysis tool," Bioinformatics, vol. 34, pp. 701–702, Feb. 2018.

[200] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," Nat. Rev. Genet., vol. 5, pp. 101–113, Feb. 2004.

[201] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," 2014.

[202] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," J Complex Netw, vol. 2, pp. 203–271, July 2014.

[203] M. Zachariou, G. Minadakis, A. Oulas, S. Afxenti, and G. M. Spyrou, "Integrating multi-source information on a single network to detect disease-related clusters of molecular mechanisms," J. Proteomics, vol. 188, pp. 15–29, Sept. 2018.

[204] D. S. Himmelstein and S. E. Baranzini, "Heterogeneous network edge prediction: A data integration approach to prioritize Disease-Associated genes," PLoS Comput. Biol., vol. 11, p. e1004259, July 2015.

[205] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas, "Ranking in interconnected multilayer networks reveals versatile nodes," Nat. Commun., vol. 6, p. 6868, Apr. 2015.

[206] A. Halu, R. J. Mondragón, P. Panzarasa, and G. Bianconi, "Multiplex PageRank," PLoS One, vol. 8, p. e78293, Oct. 2013.

[207] D. Edler, L. Bohlin, and M. Rosvall, "Mapping higher-order network flows in memory and multilayer networks with infomap," arXiv [cs.SI], vol. 10, pp. 1–23, June 2017.

[208] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, "Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems," Phys. Rev. X, vol. 5, p. 011027, Mar. 2015.

[209] S. J. Teran Hidalgo and S. Ma, "Clustering multilayer omics data using MuNCut," BMC Genomics, vol. 19, p. 198, Mar. 2018.

[210] Y. Perez-Riverol, M. Bai, F. da Veiga Leprevost, S. Squizzato, Y. M. Park, K. Haug, A. J. Carroll, D. Spalding, J. Paschall, M. Wang, N. Del-Toro, T. Ternent, P. Zhang, N. Buso, N. Bandeira, E. W. Deutsch, D. S. Campbell, R. C. Beavis, R. M. Salek, U. Sarkans, R. Petryszak, M. Keays, E. Fahy, M. Sud, S. Subramaniam, A. Barbera, R. C. Jiménez, A. I. Nesvizhskii, S.-A. Sansone, C. Steinbeck, R. Lopez, J. A. Vizcaíno, P. Ping, and H. Hermjakob, "Discovering and linking public omics data sets using the omics discovery index," Nat. Biotechnol., vol. 35, pp. 406–409, May 2017.

[211] Y. Perez-Riverol, A. Zorin, G. Dass, M.-T. Vu, P. Xu, M. Glont, J. A. Vizcaíno, A. F. Jarnuczak, R. Petryszak, P. Ping, and H. Hermjakob, "Quantifying the impact of public omics data," Nat. Commun., vol. 10, p. 3512, Aug. 2019.

[212] M. Pietzner, E. Wheeler, J. Carrasco-Zanini, A. Cortes, M. Koprulu, M. A. Wörheide, E. Oerton, J. Cook, I. D. Stewart, N. D. Kerrison, J. Luan, J. Raffler, M. Arnold, W. Arlt, S. O'Rahilly, G. Kastenmüller, E. R. Gamazon, A. D. Hingorani, R. A. Scott, N. J. Wareham, and C. Langenberg, "Mapping the proteo-genomic convergence of human diseases," Science, vol. 374, p. eabj1541, Nov. 2021.

[213] G. Eraslan, Ž. Avsec, J. Gagneur, and F. J. Theis, "Deep learning: new computational modelling techniques for genomics," 2019.

[214] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P. M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. Decaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. S. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene, Opportunities and obstacles for deep learning in biology and medicine, vol. 15. 2018.

[215] Z. Zhang, Y. Zhao, X. Liao, W. Shi, K. Li, Q. Zou, and S. Peng, "Deep learning in omics: a survey and guideline," Brief. Funct. Genomics, vol. 18, pp. 41–57, Feb. 2019.

[216] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins, "Next-Generation machine learning for biological networks," 2018.

[217] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313, pp. 504–507, July 2006.

[218] T. Gomes, S. A. Teichmann, and C. Talavera-López, "Immunology driven by Large-Scale Single-Cell sequencing," Trends Immunol., vol. 40, pp. 1011–1021, Nov. 2019.

[219] N. C. Chung, B. Mirza, H. Choi, J. Wang, D. Wang, P. Ping, and W. Wang, "Unsupervised classification of multi-omics data during cardiac remodeling using deep learning," Methods, vol. 166, pp. 66–73, Aug. 2019.

[220] X. Zhang, J. Zhang, K. Sun, X. Yang, C. Dai, and Y. Guo, "Integrated multi-omics analysis using variational autoencoders: Application to pan-cancer classification," arXiv [cs.LG], pp. 765–769, Aug. 2019.

[221] S. Webb, "Deep learning for biology," Nature, vol. 554, pp. 555–557, Feb. 2018.

[222] Z. Xiao, J. W. Locasale, and Z. Dai, "Metabolism in the tumor microenvironment: insights from single-cell analysis," Oncoimmunology, vol. 9, p. 1726556, Feb. 2020.

[223] T. Alexandrov, "Spatial metabolomics and imaging mass spectrometry in the age of artificial intelligence," Annu Rev Biomed Data Sci, vol. 3, pp. 61–87, July 2020.

[224] L. Rappez, M. Stadler, S. Triana, P. Phapale, M. Heikenwalder, and T. Alexandrov, "Spatial single-cell profiling of intracellular metabolomes in situ," bioRxiv, p. 510222, Jan. 2019.

[225] C. G. Palii, Q. Cheng, M. A. Gillespie, P. Shannon, M. Mazurczyk, G. Napolitani, N. D. Price, J. A. Ranish, E. Morrissey, D. R. Higgs, and M. Brand, "Single-Cell proteomics reveal that quantitative changes in co-expressed Lineage-Specific transcription factors determine cell fate," Cell Stem Cell, vol. 24, pp. 812–820.e5, May 2019.

[226] D. J. Burgess, "Spatial transcriptomics coming of age," Nat. Rev. Genet., vol. 20, p. 317, June 2019.

[227] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The FAIR guiding principles for scientific data management and stewardship," Sci Data, vol. 3, p. 160018, Mar. 2016.

[228] N. Toodayan, "Professor alois alzheimer (1864-1915): Lest we forget," J. Clin. Neurosci., vol. 31, pp. 47–55, Sept. 2016.

[229] A. Alzheimer, "Uber eine eigenartige erkrankung der hirnrinde," Allgemeine Zeitschrift fur Psychiatrie und Psychisch-gerichtliche Medizin, vol. 64, pp. 146–148, Jan. 1907.

[230] K. B. Rajan, J. Weuve, L. L. Barnes, E. A. McAninch, R. S. Wilson, and D. A. Evans, "Population estimate of people with clinical alzheimer's disease and mild cognitive impairment in the united states (2020-2060)," Alzheimers. Dement., vol. 17, pp. 1966–1975, Dec. 2021.

[231] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans, "Alzheimer disease in the united states (2010-2050) estimated using the 2010 census," Neurology, vol. 80, pp. 1778–1783, May 2013.

[232] "2022 alzheimer's disease facts and figures," Alzheimers. Dement., vol. 18, pp. 700–789, Apr. 2022.

[233] A. Kapasi, C. DeCarli, and J. A. Schneider, "Impact of multiple pathologies on the threshold for clinically overt dementia," Acta Neuropathol., vol. 134, pp. 171–186, Aug. 2017.

[234] E. A. Newcombe, J. Camats-Perna, M. L. Silva, N. Valmas, T. J. Huat, and R. Medeiros, "Inflammation: the link between comorbidities, genetics, and alzheimer's disease," J. Neuroinflammation, vol. 15, p. 276, Sept. 2018.

[235] T. Ayodele, E. Rogaeva, J. T. Kurup, G. Beecham, and C. Reitz, "Early-Onset alzheimer's disease: What is missing in research?," Curr. Neurol. Neurosci. Rep., vol. 21, p. 4, Jan. 2021.

[236] M. Gatz, C. A. Reynolds, L. Fratiglioni, B. Johansson, J. A. Mortimer, S. Berg, A. Fiske, and N. L. Pedersen, "Role of genes and environments for explaining alzheimer disease," Arch. Gen. Psychiatry, vol. 63, pp. 168–174, Feb. 2006.

[237] W. J. Strittmatter, K. H. Weisgraber, D. Y. Huang, L. M. Dong, G. S. Salvesen, M. Pericak-Vance, D. Schmechel, A. M. Saunders, D. Goldgaber, and A. D. Roses, "Binding of human apolipoprotein E to synthetic amyloid beta peptide: isoform-specific effects and implications for late-onset alzheimer disease," Proc. Natl. Acad. Sci. U. S. A., vol. 90, pp. 8098–8102, Sept. 1993.

[238] Y. Yamazaki, N. Zhao, T. R. Caulfield, C.-C. Liu, and G. Bu, "Apolipoprotein E and alzheimer disease: pathobiology and targeting strategies," Nat. Rev. Neurol., vol. 15, pp. 501–518, Sept. 2019.

[239] C. Bellenguez, B. Grenier-Boley, and J.-C. Lambert, "Genetics of alzheimer's disease: where we are, and where we are going," Curr. Opin. Neurobiol., vol. 61, pp. 40–48, Apr. 2020.

[240] M. X. Tang, Y. Stern, K. Marder, K. Bell, B. Gurland, R. Lantigua, H. Andrews, L. Feng, B. Tycko, and R. Mayeux, "The APOE-epsilon4 allele and the risk of alzheimer disease among african americans, whites, and hispanics," JAMA, vol. 279, pp. 751–755, Mar. 1998.

[241] C. Reitz, G. Jun, A. Naj, R. Rajbhandary, B. N. Vardarajan, L.-S. Wang, O. Valladares, C.-F. Lin, E. B. Larson, N. R. Graff-Radford, D. Evans, P. L. De Jager, P. K. Crane, J. D. Buxbaum, J. R. Murrell, T. Raj, N. Ertekin-Taner, M. Logue, C. T. Baldwin, R. C. Green, L. L. Barnes, L. B. Cantwell, M. D. Fallin, R. C. P. Go, P. Griffith, T. O. Obisesan, J. J. Manly, K. L. Lunetta, M. I. Kamboh, O. L. Lopez, D. A. Bennett, H. Hendrie, K. S. Hall, A. M. Goate, G. S. Byrd, W. A. Kukull, T. M. Foroud, J. L. Haines, L. A. Farrer, M. A. Pericak-Vance, G. D. Schellenberg, R. Mayeux, and Alzheimer Disease Genetics Consortium, "Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E e4,and the risk of late-onset alzheimer disease in african americans," JAMA, vol. 309, pp. 1483–1492, Apr. 2013.

[242] G. Chêne, A. Beiser, R. Au, S. R. Preis, P. A. Wolf, C. Dufouil, and S. Seshadri, "Gender and incidence of dementia in the framingham heart study from mid-adult life," Alzheimers. Dement., vol. 11, pp. 310–320, Mar. 2015.

[243] D. Sohn, K. Shpanskaya, J. E. Lucas, J. R. Petrella, A. J. Saykin, R. E. Tanzi, N. F. Samatova, and P. M. Doraiswamy, "Sex differences in cognitive decline in subjects with high likelihood of mild cognitive impairment due to alzheimer's disease," Sci. Rep., vol. 8, p. 7490, May 2018.

[244] R. A. Nebel, N. T. Aggarwal, L. L. Barnes, A. Gallagher, J. M. Goldstein, K. Kantarci, M. P. Mallampalli, E. C. Mormino, L. Scott, W. H. Yu, P. M. Maki, and M. M. Mielke, "Understanding the impact of sex and gender in alzheimer's disease: A call to action," Alzheimers. Dement., vol. 14, pp. 1171–1183, Sept. 2018.

[245] M. Arnold, K. Nho, A. Kueider-Paisley, T. Massaro, K. Huynh, B. Brauner, S. MahmoudianDehkordi, G. Louie, M. A. Moseley, J. W. Thompson, L. S. John-Williams, J. D. Tenenbaum, C. Blach, R. Chang, R. D. Brinton, R. Baillie, X. Han, J. Q. Trojanowski, L. M. Shaw, R. Martins, M. W. Weiner, E. Trushina, J. B. Toledo, P. J. Meikle, D. A. Bennett, J. Krumsiek, P. M. Doraiswamy, A. J. Saykin, R. Kaddurah-Daouk, and G. Kastenmüller, "Sex and APOE $\epsilon 4$ genotype modify the alzheimer's disease serum metabolome," Nat. Commun., vol. 11, p. 1148, Mar. 2020.

[246] M. J. Lennon, S. R. Makkar, J. D. Crawford, and P. S. Sachdev, "Midlife hypertension and alzheimer's disease: A systematic review and Meta-Analysis," J. Alzheimers. Dis., vol. 71, no. 1, pp. 307–316, 2019.

[247] A. L. Fitzpatrick, L. H. Kuller, O. L. Lopez, P. Diehr, E. S. O'Meara, W. T. Longstreth, Jr, and J. A. Luchsinger, "Midlife and late-life obesity and the risk of dementia: cardiovascular health study," Arch. Neurol., vol. 66, pp. 336–342, Mar. 2009.

[248] G. J. Biessels, S. Staekenborg, E. Brunner, C. Brayne, and P. Scheltens, "Risk of dementia in diabetes mellitus: a systematic review," Lancet Neurol., vol. 5, pp. 64–74, Jan. 2006.

[249] G. Livingston, J. Huntley, A. Sommerlad, D. Ames, C. Ballard, S. Banerjee, C. Brayne, A. Burns, J. Cohen-Mansfield, C. Cooper, S. G. Costafreda, A. Dias, N. Fox, L. N. Gitlin, R. Howard, H. C. Kales, M. Kivimäki, E. B. Larson, A. Ogunniyi, V. Orgeta, K. Ritchie, K. Rockwood, E. L. Sampson, Q. Samus, L. S. Schneider, G. Selbæk, L. Teri, and N. Mukadam, "Dementia prevention, intervention, and care: 2020 report of the lancet commission," Lancet, vol. 396, pp. 413–446, Aug. 2020.

[250] C. Reitz, C. Brayne, and R. Mayeux, "Epidemiology of alzheimer disease," Nat. Rev. Neurol., vol. 7, pp. 137–152, Mar. 2011.

[251] C. R. Jack, Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J. L. Molinuevo, T. Montine, C. Phelps, K. P.

Rankin, C. C. Rowe, P. Scheltens, E. Siemers, H. M. Snyder, R. Sperling, and Contributors, "NIA-AA research framework: Toward a biological definition of alzheimer's disease," Alzheimers. Dement., vol. 14, pp. 535–562, Apr. 2018.

[252] D. J. Selkoe and J. Hardy, "The amyloid hypothesis of alzheimer's disease at 25 years," EMBO Mol. Med., vol. 8, pp. 595–608, June 2016.

[253] P. Scheltens, K. Blennow, M. M. B. Breteler, B. de Strooper, G. B. Frisoni, S. Salloway, and W. M. Van der Flier, "Alzheimer's disease," Lancet, vol. 388, pp. 505–517, July 2016.

[254] R. H. Swerdlow and S. M. Khan, "A "mitochondrial cascade hypothesis" for sporadic alzheimer's disease," Med. Hypotheses, vol. 63, no. 1, pp. 8–20, 2004.

[255] S. A. Small and K. Duff, "Linking abeta and tau in late-onset alzheimer's disease: a dual pathway hypothesis," Neuron, vol. 60, pp. 534–542, Nov. 2008.

[256] P.-P. Liu, Y. Xie, X.-Y. Meng, and J.-S. Kang, "History and progress of hypotheses and clinical trials for alzheimer's disease," Signal Transduct Target Ther, vol. 4, p. 29, Aug. 2019.

[257] R. van der Kant, L. S. B. Goldstein, and R. Ossenkoppele, "Amyloid-β-independent regulators of tau pathology in alzheimer disease," Nat. Rev. Neurosci., vol. 21, pp. 21–35, Jan. 2020.

[258] H. W. Querfurth and F. M. LaFerla, "Alzheimer's disease," N. Engl. J. Med., vol. 362, pp. 329–344, Jan. 2010.

[259] D. Allan Butterfield, "Amyloid β-peptide (1-42)-induced oxidative stress and neurotoxicity: Implications for neurodegeneration in alzheimer's disease brain. a review," Free Radic. Res., vol. 36, pp. 1307–1313, Jan. 2002.

[260] M. Arbel-Ornath, E. Hudry, J. R. Boivin, T. Hashimoto, S. Takeda, K. V. Kuchibhotla, S. Hou, C. R. Lattarulo, A. M. Belcher, N. Shakerdge, P. B. Trujillo, A. Muzikansky, R. A. Betensky, B. T. Hyman, and B. J. Bacskai, "Soluble oligomeric amyloid-β induces calcium dyshomeostasis that precedes synapse loss in the living mouse brain," Mol. Neurodegener., vol. 12, p. 27, Mar. 2017.

[261] D. S. Knopman, H. Amieva, R. C. Petersen, G. Chételat, D. M. Holtzman, B. T. Hyman, R. A. Nixon, and D. T. Jones, "Alzheimer disease," Nat Rev Dis Primers, vol. 7, p. 33, May 2021.

[262] R. Ossenkoppele, D. R. Schonhaut, M. Schöll, S. N. Lockhart, N. Ayakta, S. L. Baker, J. P. O'Neil, M. Janabi, A. Lazaris, A. Cantwell, J. Vogel, M. Santos, Z. A. Miller, B. M. Bettcher, K. A. Vossel, J. H. Kramer, M. L. Gorno-Tempini, B. L. Miller, W. J. Jagust, and G. D. Rabinovici, "Tau PET patterns mirror clinical and neuroanatomical variability in alzheimer's disease," Brain, vol. 139, pp. 1551–1567, May 2016.

[263] R. Batra, M. Arnold, M. A. Wörheide, M. Allen, X. Wang, C. Blach, A. I. Levey, N. T. Seyfried, N. Ertekin-Taner, D. A. Bennett, G. Kastenmüller, R. F. Kaddurah-Daouk, J. Krumsiek, and Alzheimer's Disease Metabolomics Consortium (ADMC), "The landscape of metabolic brain alterations in alzheimer's disease," Alzheimers. Dement., July 2022.

[264] N. Maphis, G. Xu, O. N. Kokiko-Cochran, S. Jiang, A. Cardona, R. M. Ransohoff, B. T. Lamb, and K. Bhaskar, "Reactive microglia drive tau pathology and contribute to the spreading of pathological tau in the brain," Brain, vol. 138, pp. 1738–1755, June 2015.

[265] Y. Shi, K. Yamada, S. A. Liddelow, S. T. Smith, L. Zhao, W. Luo, R. M. Tsai, S. Spina, L. T. Grinberg, J. C. Rojas, G. Gallardo, K. Wang, J. Roh, G. Robinson, M. B. Finn, H. Jiang, P. M. Sullivan, C. Baufeld, M. W. Wood, C. Sutphen, L. McCue, C. Xiong, J. L. Del-Aguila, J. C. Morris, C. Cruchaga, Alzheimer's Disease Neuroimaging Initiative, A. M. Fagan, B. L. Miller, A. L. Boxer, W. W. Seeley, O. Butovsky, B. A. Barres, S. M. Paul, and D. M. Holtzman, "ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy," Nature, vol. 549, pp. 523–527, Sept. 2017.

[266] C. M. Henstridge, B. T. Hyman, and T. L. Spires-Jones, "Beyond the neuron-cellular interactions early in alzheimer disease pathogenesis," Nat. Rev. Neurosci., vol. 20, pp. 94–108, Feb. 2019.

[267] R. D. Terry, E. Masliah, D. P. Salmon, N. Butters, R. DeTeresa, R. Hill, L. A. Hansen, and R. Katzman, "Physical basis of cognitive alterations in alzheimer's disease: synapse loss is the major correlate of cognitive impairment," Ann. Neurol., vol. 30, pp. 572–580, Oct. 1991.

[268] L. Zhou, J. McInnes, K. Wierda, M. Holt, A. G. Herrmann, R. J. Jackson, Y.-C. Wang, J. Swerts, J. Beyens, K. Miskiewicz, S. Vilain, I. Dewachter, D. Moechars, B. De Strooper, T. L. Spires-Jones, J. De Wit, and P. Verstreken, "Tau association with synaptic vesicles causes presynaptic dysfunction," Nat. Commun., vol. 8, p. 15295, May 2017.

[269] T. L. Spires-Jones and B. T. Hyman, "The intersection of amyloid beta and tau at synapses in alzheimer's disease," Neuron, vol. 82, pp. 756–771, May 2014.

[270] G. M. Shankar, B. L. Bloodgood, M. Townsend, D. M. Walsh, D. J. Selkoe, and B. L. Sabatini, "Natural oligomers of the alzheimer amyloid-beta protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway," J. Neurosci., vol. 27, pp. 2866–2875, Mar. 2007.

[271] A. A. George, J. M. Vieira, C. Xavier-Jackson, M. T. Gee, J. R. Cirrito, H. A. Bimonte-Nelson, M. R. Picciotto, R. J. Lukas, and P. Whiteaker, "Implications of oligomeric Amyloid-Beta (oAβ42) signaling through $\alpha7\beta2$-Nicotinic acetylcholine receptors (nAChRs) on basal forebrain cholinergic neuronal intrinsic excitability and cognitive decline," J. Neurosci., vol. 41, pp. 555–575, Jan. 2021.

[272] B. W. Kunkle, B. Grenier-Boley, R. Sims, J. C. Bis, V. Damotte, A. C. Naj, A. Boland, M. Vronskaya, S. J. van der Lee, A. Amlie-Wolf, C. Bellenguez, A. Frizatti, V. Chouraki, E. R. Martin, K. Sleegers, N. Badarinarayan, J. Jakobsdottir, K. L. Hamilton-Nelson, S. Moreno-Grau, R. Olaso, R. Raybould, Y. Chen, A. B. Kuzma, M. Hiltunen, T. Morgan, S. Ahmad, B. N. Vardarajan, J. Epelbaum, P. Hoffmann, M. Boada, G. W. Beecham, J.-G. Garnier, D. Harold, A. L. Fitzpatrick, O. Valladares, M.-L. Moutet, A. Gerrish, A. V. Smith, L. Qu, D. Bacq, N. Denning, X. Jian, Y. Zhao, M. Del Zompo, N. C. Fox, S.-H. Choi, I. Mateo, J. T. Hughes, H. H. Adams, J. Malamon, F. Sanchez-Garcia, Y. Patel, J. A. Brody, B. A. Dombroski, M. C. D. Naranjo, M. Daniilidou, G. Eiriksdottir, S. Mukherjee, D. Wallon, J. Uphill, T. Aspelund, L. B. Cantwell, F. Garzia, D. Galimberti, E. Hofer, M. Butkiewicz, B. Fin, E. Scarpini, C. Sarnowski, W. S. Bush, S. Meslage, J. Kornhuber, C. C. White, Y. Song, R. C. Barber, S. Engelborghs, S. Sordon, D. Voijnovic, P. M. Adams, R. Vandenberghe, M. Mayhaus, L. A. Cupples, M. S. Albert, P. P. De Deyn, W. Gu, J. J. Himali, D. Beekly, A. Squassina, A. M. Hartmann, A. Orellana, D. Blacker, E. Rodriguez-Rodriguez, S. Lovestone, M. E. Garcia, R. S. Doody, C. Munoz-Fernadez, R. Sussams, H. Lin, T. J. Fairchild, Y. A. Benito, C. Holmes, H. Karamujić-Čomić, M. P. Frosch, H. Thonberg, W. Maier, G. Roshchupkin, B. Ghetti, V. Giedraitis, A. Kawalia, S. Li, R. M. Huebinger, L. Kilander, S. Moebus, I. Hernández, M. I. Kamboh, R. Brundin, J. Turton, Q. Yang, M. J. Katz, L. Concari, J. Lord, A. S. Beiser, C. D. Keene, S. Helisalmi, I. Kloszewska, W. A. Kukull, A. M. Koivisto, A. Lynch, L. Tarraga, E. B. Larson, A. Haapasalo, B. Lawlor, T. H. Mosley, R. B. Lipton, V. Solfrizzi, M. Gill, W. T. Longstreth, Jr, T. J. Montine, V. Frisardi, M. Diez-Fairen, F. Rivadeneira, R. C. Petersen, V. Deramecourt, I. Alvarez, F. Salani, A. Ciaramella, E. Boerwinkle, E. M. Reiman, N. Fievet, J. I. Rotter, J. S. Reisch, O. Hanon, C. Cupidi, A. G. Andre Uitterlinden, D. R. Royall, C. Dufouil, R. G. Maletta, I. de Rojas, M. Sano, A. Brice, R. Cecchetti, P. S. George-Hyslop, K. Ritchie, M. Tsolaki, D. W. Tsuang, B. Dubois, D. Craig, C.-K. Wu, H. Soininen, D. Avramidou, R. L. Albin, L. Fratiglioni, A. Germanou, L. G. Apostolova, L. Keller, M. Koutroumani, S. E. Arnold, F. Panza, O. Gkatzima, S. Asthana, D. Hannequin, P. Whitehead, C. S. Atwood, P. Caffarra, H. Hampel, I. Quintela, Á. Carracedo, L. Lannfelt, D. C. Rubinsztein, L. L. Barnes, F. Pasquier, L. Frölich, S. Barral, B. McGuinness, T. G. Beach, J. A. Johnston, J. T. Becker, P. Passmore, E. H. Bigio, J. M. Schott, T. D. Bird, J. D. Warren, B. F. Boeve, M. K. Lupton, J. D. Bowen, P. Proitsi,

A. Boxer, J. F. Powell, J. R. Burke, J. S. K. Kauwe, J. M. Burns, M. Mancuso, J. D. Buxbaum, U. Bonuccelli, N. J. Cairns, A. McQuillin, C. Cao, G. Livingston, C. S. Carlson, N. J. Bass, C. M. Carlsson, J. Hardy, R. M. Carney, J. Bras, M. M. Carrasquillo, R. Guerreiro, M. Allen, H. C. Chui, E. Fisher, C. Masullo, E. A. Crocco, C. DeCarli, G. Bisceglio, M. Dick, L. Ma, R. Duara, N. R. Graff-Radford, D. A. Evans, A. Hodges, K. M. Faber, M. Scherer, K. B. Fallon, M. Riemenschneider, D. W. Fardo, R. Heun, M. R. Farlow, H. Kölsch, S. Ferris, M. Leber, T. M. Foroud, I. Heuser, D. R. Galasko, I. Giegling, M. Gearing, M. Hüll, D. H. Geschwind, J. R. Gilbert, J. Morris, R. C. Green, K. Mayo, J. H. Growdon, T. Feulner, R. L. Hamilton, L. E. Harrell, D. Drichel, L. S. Honig, T. D. Cushion, M. J. Huentelman, P. Hollingworth, C. M. Hulette, B. T. Hyman, R. Marshall, G. P. Jarvik, A. Meggy, E. Abner, G. E. Menzies, L.-W. Jin, G. Leonenko, L. M. Real, G. R. Jun, C. T. Baldwin, D. Grozeva, A. Karydas, G. Russo, J. A. Kaye, R. Kim, F. Jessen, N. W. Kowall, B. Vellas, J. H. Kramer, E. Vardy, F. M. LaFerla, K.-H. Jöckel, J. J. Lah, M. Dichgans, J. B. Leverenz, D. Mann, A. I. Levey, S. Pickering-Brown, A. P. Lieberman, N. Klopp, K. L. Lunetta, H.-E. Wichmann, C. G. Lyketsos, K. Morgan, D. C. Marson, K. Brown, F. Martiniuk, C. Medway, D. C. Mash, M. M. Nöthen, E. Masliah, N. M. Hooper, W. C. McCormick, A. Daniele, S. M. McCurry, A. Bayer, A. N. McDavid, J. Gallacher, A. C. McKee, H. van den Bussche, M. Mesulam, C. Brayne, B. L. Miller, S. Riedel-Heller, C. A. Miller, J. W. Miller, A. Al-Chalabi, J. C. Morris, C. E. Shaw, A. J. Myers, J. Wiltfang, S. O'Bryant, J. M. Olichney, V. Alvarez, J. E. Parisi, A. B. Singleton, H. L. Paulson, J. Collinge, W. R. Perry, S. Mead, E. Peskind, D. H. Cribbs, M. Rossor, A. Pierce, N. S. Ryan, W. W. Poon, B. Nacmias, H. Potter, S. Sorbi, J. F. Quinn, E. Sacchinelli, A. Raj, G. Spalletta, M. Raskind, C. Caltagirone, P. Bossù, M. D. Orfei, B. Reisberg, R. Clarke, C. Reitz, A. D. Smith, J. M. Ringman, D. Warden, E. D. Roberson, G. Wilcock, E. Rogaeva, A. C. Bruni, H. J. Rosen, M. Gallo, R. N. Rosenberg, Y. Ben-Shlomo, M. A. Sager, P. Mecocci, A. J. Saykin, P. Pastor, M. L. Cuccaro, J. M. Vance, J. A. Schneider, L. S. Schneider, S. Slifer, W. W. Seeley, A. G. Smith, J. A. Sonnen, S. Spina, R. A. Stern, R. H. Swerdlow, M. Tang, R. E. Tanzi, J. Q. Trojanowski, J. C. Troncoso, V. M. Van Deerlin, L. J. Van Eldik, H. V. Vinters, J. P. Vonsattel, S. Weintraub, K. A. Welsh-Bohmer, K. C. Wilhelmsen, J. Williamson, T. S. Wingo, R. L. Woltjer, C. B. Wright, C.-E. Yu, L. Yu, Y. Saba, A. Pilotto, M. J. Bullido, O. Peters, P. K. Crane, D. Bennett, P. Bosco, E. Coto, V. Boccardi, P. L. De Jager, A. Lleo, N. Warner, O. L. Lopez, M. Ingelsson, P. Deloukas, C. Cruchaga, C. Graff, R. Gwilliam, M. Fornage, A. M. Goate, P. Sanchez-Juan, P. G. Kehoe, N. Amin, N. Ertekin-Taner, C. Berr, S. Debette, S. Love, L. J. Launer, S. G. Younkin, J.-F. Dartigues, C. Corcoran, M. A. Ikram, D. W. Dickson, G. Nicolas, D. Campion, J. Tschanz, H. Schmidt, H. Hakonarson, J. Clarimon, R. Munger, R. Schmidt, L. A. Farrer, C. Van Broeckhoven, M. C O'Donovan, A. L. DeStefano, L. Jones, J. L. Haines, J.-F. Deleuze, M. J. Owen, V. Gudnason, R. Mayeux, V. Escott-Price, B. M. Psaty, A. Ramirez, L.-S. Wang, A. Ruiz, C. M. van Duijn, P. A. Holmans, S. Seshadri, J. Williams, P. Amouyel, G. D. Schellenberg, J.-C. Lambert, M. A. Pericak-Vance, Alzheimer Disease Genetics Consortium (ADGC),, European Alzheimer's Disease Initiative (EADI),, Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE),, and Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES),, "Genetic meta-analysis of diagnosed alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing," Nat. Genet., vol. 51, pp. 414–430, Mar. 2019.

[273] C. Bellenguez, F. Küçükali, I. E. Jansen, L. Kleineidam, S. Moreno-Grau, N. Amin, A. C. Naj, R. Campos-Martin, B. Grenier-Boley, V. Andrade, P. A. Holmans, A. Boland, V. Damotte, S. J. van der Lee, M. R. Costa, T. Kuulasmaa, Q. Yang, I. de Rojas, J. C. Bis, A. Yaqub, I. Prokic, J. Chapuis, S. Ahmad, V. Giedraitis, D. Aarsland, P. Garcia-Gonzalez, C. Abdelnour, E. Alarcón-Martín, D. Alcolea, M. Alegret, I. Alvarez, V. Álvarez, N. J. Armstrong, A. Tsolaki, C. Antúnez, I. Appollonio, M. Arcaro, S. Archetti, A. A. Pastor, B. Arosio, L. Athanasiu, H. Bailly, N. Banaj, M. Baquero, S. Barral, A. Beiser, A. B. Pastor, J. E. Below, P. Benchek, L. Benussi, C. Berr, C. Besse, V. Bessi, G. Binetti, A. Bizarro, R. Blesa, M. Boada, E. Boerwinkle, B. Borroni, S. Boschi, P. Bossù,

G. Bråthen, J. Bressler, C. Bresner, H. Brodaty, K. J. Brookes, L. I. Brusco, D. Buiza-Rueda, K. Bûrger, V. Burholt, W. S. Bush, M. Calero, L. B. Cantwell, G. Chene, J. Chung, M. L. Cuccaro, Á. Carracedo, R. Cecchetti, L. Cervera-Carles, C. Charbonnier, H.-H. Chen, C. Chillotti, S. Ciccone, J. A. H. R. Claassen, C. Clark, E. Conti, A. Corma-Gómez, E. Costantini, C. Custodero, D. Daian, M. C. Dalmasso, A. Daniele, E. Dardiotis, J.-F. Dartigues, P. P. de Deyn, K. de Paiva Lopes, L. D. de Witte, S. Debette, J. Deckert, T. Del Ser, N. Denning, A. DeStefano, M. Dichgans, J. Diehl-Schmid, M. Diez-Fairen, P. D. Rossi, S. Djurovic, E. Duron, E. Düzel, C. Dufouil, G. Eiriksdottir, S. Engelborghs, V. Escott-Price, A. Espinosa, M. Ewers, K. M. Faber, T. Fabrizio, S. F. Nielsen, D. W. Fardo, L. Farotti, C. Fenoglio, M. Fernández-Fuertes, R. Ferrari, C. B. Ferreira, E. Ferri, B. Fin, P. Fischer, T. Fladby, K. Fließbach, B. Fongang, M. Fornage, J. Fortea, T. M. Foroud, S. Fostinelli, N. C. Fox, E. Franco-Macías, M. J. Bullido, A. Frank-García, L. Froelich, B. Fulton-Howard, D. Galimberti, J. M. García-Alberca, P. García-González, S. Garcia-Madrona, G. Garcia-Ribas, R. Ghidoni, I. Giegling, G. Giorgio, A. M. Goate, O. Goldhardt, D. Gomez-Fonseca, A. González-Pérez, C. Graff, G. Grande, E. Green, T. Grimmer, E. Grünblatt, M. Grunin, V. Gudnason, T. Guetta-Baranes, A. Haapasalo, G. Hadjigeorgiou, J. L. Haines, K. L. Hamilton-Nelson, H. Hampel, O. Hanon, J. Hardy, A. M. Hartmann, L. Hausner, J. Harwood, S. Heilmann-Heimbach, S. Helisalmi, M. T. Heneka, I. Hernández, M. J. Herrmann, P. Hoffmann, C. Holmes, H. Holstege, R. H. Vilas, M. Hulsman, J. Humphrey, G. J. Biessels, X. Jian, C. Johansson, G. R. Jun, Y. Kastumata, J. Kauwe, P. G. Kehoe, L. Kilander, A. K. Ståhlbom, M. Kivipelto, A. Koivisto, J. Kornhuber, M. H. Kosmidis, W. A. Kukull, P. P. Kuksa, B. W. Kunkle, A. B. Kuzma, C. Lage, E. J. Laukka, L. Launer, A. Lauria, C.-Y. Lee, J. Lehtisalo, O. Lerch, A. Lleó, W. Longstreth, Jr, O. Lopez, A. L. de Munain, S. Love, M. Löwemark, L. Luckcuck, K. L. Lunetta, Y. Ma, J. Macías, C. A. MacLeod, W. Maier, F. Mangialasche, M. Spallazzi, M. Marquié, R. Marshall, E. R. Martin, A. M. Montes, C. M. Rodríguez, C. Masullo, R. Mayeux, S. Mead, P. Mecocci, M. Medina, A. Meggy, S. Mehrabian, S. Mendoza, M. Menéndez-González, P. Mir, S. Moebus, M. Mol, L. Molina-Porcel, L. Montrreal, L. Morelli, F. Moreno, K. Morgan, T. Mosley, M. M. Nöthen, C. Muchnik, S. Mukherjee, B. Nacmias, T. Ngandu, G. Nicolas, B. G. Nordestgaard, R. Olaso, A. Orellana, M. Orsini, G. Ortega, A. Padovani, C. Paolo, G. Papenberg, L. Parnetti, F. Pasquier, P. Pastor, G. Peloso, A. Pérez-Cordón, J. Pérez-Tur, P. Pericard, O. Peters, Y. A. L. Pijnenburg, J. A. Pineda, G. Piñol-Ripoll, C. Pisanu, T. Polak, J. Popp, D. Posthuma, J. Priller, R. Puerta, O. Quenez, I. Quintela, J. Q. Thomassen, A. Rábano, I. Rainero, F. Rajabli, I. Ramakers, L. M. Real, M. J. T. Reinders, C. Reitz, D. Reyes-Dumeyer, P. Ridge, S. Riedel-Heller, P. Riederer, N. Roberto, E. Rodriguez-Rodriguez, A. Rongve, I. R. Allende, M. Rosende-Roca, J. L. Royo, E. Rubino, D. Rujescu, M. E. Sáez, P. Sakka, I. Saltvedt, Á. Sanabria, M. B. Sánchez-Arjona, F. Sanchez-Garcia, P. S. Juan, R. Sánchez-Valle, S. B. Sando, C. Sarnowski, C. L. Satizabal, M. Scamosci, N. Scarmeas, E. Scarpini, P. Scheltens, N. Scherbaum, M. Scherer, M. Schmid, A. Schneider, J. M. Schott, G. Selbæk, D. Seripa, M. Serrano, J. Sha, A. A. Shadrin, O. Skrobot, S. Slifer, G. J. L. Snijders, H. Soininen, V. Solfrizzi, A. Solomon, Y. Song, S. Sorbi, O. Sotolongo-Grau, G. Spalletta, A. Spottke, A. Squassina, E. Stordal, J. P. Tartan, L. Tárraga, N. Tesí, A. Thalamuthu, T. Thomas, G. Tosto, L. Traykov, L. Tremolizzo, A. Tybjærg-Hansen, A. Uitterlinden, A. Ullgren, I. Ulstein, S. Valero, O. Valladares, C. Van Broeckhoven, J. Vance, B. N. Vardarajan, A. van der Lugt, J. Van Dongen, J. van Rooij, J. van Swieten, R. Vandenberghe, F. Verhey, J.-S. Vidal, J. Vogelgsang, M. Vyhnalek, M. Wagner, D. Wallon, L.-S. Wang, R. Wang, L. Weinhold, J. Wiltfang, G. Windle, B. Woods, M. Yannakoulia, H. Zare, Y. Zhao, X. Zhang, C. Zhu, M. Zulaica, EADB, GR@ACE, DEGESCO, EADI, GERAD, Demgene, FinnGen, ADGC, CHARGE, L. A. Farrer, B. M. Psaty, M. Ghanbari, T. Raj, P. Sachdev, K. Mather, F. Jessen, M. A. Ikram, A. de Mendonça, J. Hort, M. Tsolaki, M. A. Pericak-Vance, P. Amouyel, J. Williams, R. Frikke-Schmidt, J. Clarimon, J.-F. Deleuze, G. Rossi, S. Seshadri, O. A. Andreassen, M. Ingelsson, M. Hiltunen, K. Sleegers, G. D. Schellenberg, C. M. van Duijn, R. Sims, W. M. van der Flier, A. Ruiz, A. Ramirez, and J.-C. Lambert, "New insights into the genetic etiology of alzheimer's disease and related dementias," Nat. Genet., vol. 54, pp. 412–436, Apr. 2022.

[274] I. E. Jansen, J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, S. Steinberg, J. Sealock, I. K. Karlsson, S. Hägg, L. Athanasiu, N. Voyle, P. Proitsi, A. Witoelar, S. Stringer, D. Aarsland, I. S. Almdahl, F. Andersen, S. Bergh, F. Bettella, S. Bjornsson, A. Brækhus, G. Bråthen, C. de Leeuw, R. S. Desikan, S. Djurovic, L. Dumitrescu, T. Fladby, T. J. Hohman, P. V. Jonsson, S. J. Kiddle, A. Rongve, I. Saltvedt, S. B. Sando, G. Selbæk, M. Shoai, N. G. Skene, J. Snaedal, E. Stordal, I. D. Ulstein, Y. Wang, L. R. White, J. Hardy, J. Hjerling-Leffler, P. F. Sullivan, W. M. van der Flier, R. Dobson, L. K. Davis, H. Stefansson, K. Stefansson, N. L. Pedersen, S. Ripke, O. A. Andreassen, and D. Posthuma, "Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer's disease risk," Nat. Genet., vol. 51, pp. 404–413, Mar. 2019.

[275] R. E. Marioni, S. E. Harris, Q. Zhang, A. F. McRae, S. P. Hagenaars, W. D. Hill, G. Davies, C. W. Ritchie, C. R. Gale, J. M. Starr, A. M. Goate, D. J. Porteous, J. Yang, K. L. Evans, I. J. Deary, N. R. Wray, and P. M. Visscher, "GWAS on family history of alzheimer's disease," Transl. Psychiatry, vol. 8, p. 99, May 2018.

[276] P. Baloni, C. C. Funk, J. Yan, J. T. Yurkovich, A. Kueider-Paisley, K. Nho, A. Heinken, W. Jia, S. Mahmoudiandehkordi, G. Louie, A. J. Saykin, M. Arnold, G. Kastenmüller, W. J. Griffiths, I. Thiele, Alzheimer's Disease Metabolomics Consortium, R. Kaddurah-Daouk, and N. D. Price, "Metabolic network analysis reveals altered bile acid synthesis and metabolism in alzheimer's disease," Cell Rep Med, vol. 1, p. 100138, Nov. 2020.

[277] P. Baloni, M. Arnold, L. Buitrago, K. Nho, H. Moreno, K. Huynh, B. Brauner, G. Louie, A. Kueider-Paisley, K. Suhre, A. J. Saykin, K. Ekroos, P. J. Meikle, L. Hood, N. D. Price, Alzheimer's Disease Metabolomics Consortium, P. M. Doraiswamy, C. C. Funk, A. I. Hernández, G. Kastenmüller, R. Baillie, X. Han, and R. Kaddurah-Daouk, "Multi-Omic analyses characterize the ceramide/sphingomyelin pathway as a therapeutic target in alzheimer's disease," Commun Biol, vol. 5, p. 1074, Oct. 2022.

[278] J. C. Lambert, C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, A. L. DeStafano, J. C. Bis, G. W. Beecham, B. Grenier-Boley, G. Russo, T. A. Thorton-Wells, N. Jones, A. V. Smith, V. Chouraki, C. Thomas, M. A. Ikram, D. Zelenika, B. N. Vardarajan, Y. Kamatani, C. F. Lin, A. Gerrish, H. Schmidt, B. Kunkle, M. L. Dunstan, A. Ruiz, M. T. Bihoreau, S. H. Choi, C. Reitz, F. Pasquier, C. Cruchaga, D. Craig, N. Amin, C. Berr, O. L. Lopez, P. L. De Jager, V. Deramecourt, J. A. Johnston, D. Evans, S. Lovestone, L. Letenneur, F. J. Morón, D. C. Rubinsztein, G. Eiriksdottir, K. Sleegers, A. M. Goate, N. Fiévet, M. W. Huentelman, M. Gill, K. Brown, M. I. Kamboh, L. Keller, P. Barberger-Gateau, B. McGuiness, E. B. Larson, R. Green, A. J. Myers, C. Dufouil, S. Todd, D. Wallon, S. Love, E. Rogaeva, J. Gallacher, P. St George-Hyslop, J. Clarimon, A. Lleo, A. Bayer, D. W. Tsuang, L. Yu, M. Tsolaki, P. Bossù, G. Spalletta, P. Proitsi, J. Collinge, S. Sorbi, F. Sanchez-Garcia, N. C. Fox, J. Hardy, M. C. Deniz Naranjo, P. Bosco, R. Clarke, C. Brayne, D. Galimberti, M. Mancuso, F. Matthews, European Alzheimer's Disease Initiative (EADI), Genetic and Environmental Risk in Alzheimer's Disease, Alzheimer's Disease Genetic Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology, S. Moebus, P. Mecocci, M. Del Zompo, W. Maier, H. Hampel, A. Pilotto, M. Bullido, F. Panza, P. Caffarra, B. Nacmias, J. R. Gilbert, M. Mayhaus, L. Lannefelt, H. Hakonarson, S. Pichler, M. M. Carrasquillo, M. Ingelsson, D. Beekly, V. Alvarez, F. Zou, O. Valladares, S. G. Younkin, E. Coto, K. L. Hamilton-Nelson, W. Gu, C. Razquin, P. Pastor, I. Mateo, M. J. Owen, K. M. Faber, P. V. Jonsson, O. Combarros, M. C. O'Donovan, L. B. Cantwell, H. Soininen, D. Blacker, S. Mead, T. H. Mosley, Jr, D. A. Bennett, T. B. Harris, L. Fratiglioni, C. Holmes, R. F. de Bruijn, P. Passmore, T. J. Montine, K. Bettens, J. I. Rotter, A. Brice, K. Morgan, T. M. Foroud, W. A. Kukull, D. Hannequin, J. F. Powell, M. A. Nalls, K. Ritchie, K. L. Lunetta, J. S. Kauwe, E. Boerwinkle, M. Riemenschneider, M. Boada, M. Hiltuenen, E. R. Martin, R. Schmidt, D. Rujescu, L. S. Wang, J. F. Dartigues, R. Mayeux, C. Tzourio, A. Hofman, M. M. Nöthen, C. Graff, B. M. Psaty, L. Jones, J. L. Haines, P. A. Holmans, M. Lathrop, M. A. Pericak-Vance, L. J. Launer, L. A. Farrer, C. M. van Duijn, C. Van Broeckhoven, V. Moskvina, S. Seshadri, J. Williams, G. D.

Schellenberg, and P. Amouyel, "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for alzheimer's disease," Nat. Genet., vol. 45, pp. 1452–1458, Dec. 2013.

[279] F. M. Feringa and R. van der Kant, "Cholesterol and alzheimer's disease; from risk genes to pathological effects," Front. Aging Neurosci., vol. 13, p. 690372, June 2021.

[280] K. Nho, A. Kueider-Paisley, S. MahmoudianDehkordi, M. Arnold, S. L. Risacher, G. Louie, C. Blach, R. Baillie, X. Han, G. Kastenmüller, W. Jia, G. Xie, S. Ahmad, T. Hankemeier, C. M. van Duijn, J. Q. Trojanowski, L. M. Shaw, M. W. Weiner, P. M. Doraiswamy, A. J. Saykin, R. Kaddurah-Daouk, and Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium, "Altered bile acid profile in mild cognitive impairment and alzheimer's disease: Relationship to neuroimaging and CSF biomarkers," Alzheimers. Dement., vol. 15, pp. 232–244, Feb. 2019.

[281] S. MahmoudianDehkordi, M. Arnold, K. Nho, S. Ahmad, W. Jia, G. Xie, G. Louie, A. Kueider-Paisley, M. A. Moseley, J. W. Thompson, L. St John Williams, J. D. Tenenbaum, C. Blach, R. Baillie, X. Han, S. Bhattacharyya, J. B. Toledo, S. Schafferer, S. Klein, T. Koal, S. L. Risacher, M. A. Kling, A. Motsinger-Reif, D. M. Rotroff, J. Jack, T. Hankemeier, D. A. Bennett, P. L. De Jager, J. Q. Trojanowski, L. M. Shaw, M. W. Weiner, P. M. Doraiswamy, C. M. van Duijn, A. J. Saykin, G. Kastenmüller, R. Kaddurah-Daouk, and Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium, "Altered bile acid profile associates with cognitive impairment in alzheimer's disease-an emerging role for gut microbiome," Alzheimers. Dement., vol. 15, pp. 76–92, Jan. 2019.

[282] F. L. Heppner, R. M. Ransohoff, and B. Becher, "Immune attack: the role of inflammation in alzheimer disease," Nat. Rev. Neurosci., vol. 16, pp. 358–372, June 2015.

[283] R. Guerreiro, A. Wojtas, J. Bras, M. Carrasquillo, E. Rogaeva, E. Majounie, C. Cruchaga, C. Sassi, J. S. K. Kauwe, S. Younkin, L. Hazrati, J. Collinge, J. Pocock, T. Lashley, J. Williams, J.-C. Lambert, P. Amouyel, A. Goate, R. Rademakers, K. Morgan, J. Powell, P. St George-Hyslop, A. Singleton, J. Hardy, and Alzheimer Genetic Analysis Group, "TREM2 variants in alzheimer's disease," N. Engl. J. Med., vol. 368, pp. 117–127, Jan. 2013.

[284] N. T. Seyfried, E. B. Dammer, V. Swarup, D. Nandakumar, D. M. Duong, L. Yin, Q. Deng, T. Nguyen, C. M. Hales, T. Wingo, J. Glass, M. Gearing, M. Thambisetty, J. C. Troncoso, D. H. Geschwind, J. J. Lah, and A. I. Levey, "A multi-network approach identifies Protein-Specific co-expression in asymptomatic and symptomatic alzheimer's disease," Cell Syst, vol. 4, pp. 60–72.e4, Jan. 2017.

[285] B. Zhang, C. Gaiteri, L.-G. Bodea, Z. Wang, J. McElwee, A. A. Podtelezhnikov, C. Zhang, T. Xie, L. Tran, R. Dobrin, E. Fluder, B. Clurman, S. Melquist, M. Narayanan, C. Suver, H. Shah, M. Mahajan, T. Gillis, J. Mysore, M. E. MacDonald, J. R. Lamb, D. A. Bennett, C. Molony, D. J. Stone, V. Gudnason, A. J. Myers, E. E. Schadt, H. Neumann, J. Zhu, and V. Emilsson, "Integrated systems approach identifies genetic nodes and networks in late-onset alzheimer's disease," Cell, vol. 153, pp. 707–720, Apr. 2013.

[286] H. Fillit, W. H. Ding, L. Buee, J. Kalman, L. Altstiel, B. Lawlor, and G. Wolf-Klein, "Elevated circulating tumor necrosis factor levels in alzheimer's disease," Neurosci. Lett., vol. 129, pp. 318–320, Aug. 1991.

[287] V. Haage and P. L. De Jager, "Neuroimmune contributions to alzheimer's disease: a focus on human data," Mol. Psychiatry, vol. 27, pp. 3164–3181, Aug. 2022.

[288] D. A. Galloway, A. E. M. Phillips, D. R. J. Owen, and C. S. Moore, "Phagocytosis in the brain: Homeostasis and disease," Front. Immunol., vol. 10, p. 790, Apr. 2019.

[289] H. Keren-Shaul, A. Spinrad, A. Weiner, O. Matcovitch-Natan, R. Dvir-Szternfeld, T. K. Ulland, E. David, K. Baruch, D. Lara-Astaiso, B. Toth, S. Itzkovitz, M. Colonna, M. Schwartz, and I. Amit, "A unique microglia type associated with restricting development of alzheimer's disease," Cell, vol. 169, pp. 1276–1290.e17, June 2017.

[290] S. Krasemann, C. Madore, R. Cialic, C. Baufeld, N. Calcagno, R. El Fatimy, L. Beckers, E. O'Loughlin, Y. Xu, Z. Fanek, D. J. Greco, S. T. Smith, G. Tweet, Z. Humulock, T. Zrzavy, P. Conde-Sanroman, M. Gacias, Z. Weng, H. Chen, E. Tjon, F. Mazaheri, K. Hartmann, A. Madi, J. D. Ulrich, M. Glatzel, A. Worthmann, J. Heeren, B. Budnik, C. Lemere, T. Ikezu, F. L. Heppner, V. Litvak, D. M. Holtzman, H. Lassmann, H. L. Weiner, J. Ochando, C. Haass, and O. Butovsky, "The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases," Immunity, vol. 47, pp. 566–581.e9, Sept. 2017.

[291] L. Vermunt, S. A. M. Sikkes, A. van den Hout, R. Handels, I. Bos, W. M. van der Flier, S. Kern, P.-J. Ousset, P. Maruff, I. Skoog, F. R. J. Verhey, Y. Freund-Levi, M. Tsolaki, Å. K. Wallin, M. Olde Rikkert, H. Soininen, L. Spiru, H. Zetterberg, K. Blennow, P. Scheltens, G. Muniz-Terrera, P. J. Visser, Alzheimer Disease Neuroimaging Initiative, AIBL Research Group, and ICTUS/DSA study groups, "Duration of preclinical, prodromal, and dementia stages of alzheimer's disease in relation to age, sex, and APOE genotype," Alzheimers. Dement., vol. 15, pp. 888–898, July 2019.

[292] R. C. Mohs, D. Knopman, R. C. Petersen, S. H. Ferris, C. Ernesto, M. Grundman, M. Sano, L. Bieliauskas, D. Geldmacher, C. Clark, and L. J. Thal, "Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the alzheimer's disease assessment scale that broaden its scope. the alzheimer's disease cooperative study," Alzheimer Dis. Assoc. Disord., vol. 11 Suppl 2, pp. S13–21, 1997.

[293] J. C. Morris, "The clinical dementia rating (CDR): current version and scoring rules," Neurology, vol. 43, pp. 2412–2414, Nov. 1993.

[294] H. Braak and E. Braak, "Neuropathological stageing of alzheimer-related changes," Acta Neuropathol., vol. 82, no. 4, pp. 239–259, 1991.

[295] S. S. Mirra, A. Heyman, D. McKeel, S. M. Sumi, B. J. Crain, L. M. Brownlee, F. S. Vogel, J. P. Hughes, G. van Belle, and L. Berg, "The consortium to establish a registry for alzheimer's disease (CERAD). part II. standardization of the neuropathologic assessment of alzheimer's disease," Neurology, vol. 41, pp. 479–486, Apr. 1991.

[296] B. T. Hyman and J. Q. Trojanowski, "Consensus recommendations for the postmortem diagnosis of alzheimer disease from the national institute on aging and the reagan institute working group on diagnostic criteria for the neuropathological assessment of alzheimer disease," J. Neuropathol. Exp. Neurol., vol. 56, pp. 1095–1097, Oct. 1997.

[297] B. T. Hyman, C. H. Phelps, T. G. Beach, E. H. Bigio, N. J. Cairns, M. C. Carrillo, D. W. Dickson, C. Duyckaerts, M. P. Frosch, E. Masliah, S. S. Mirra, P. T. Nelson, J. A. Schneider, D. R. Thal, B. Thies, J. Q. Trojanowski, H. V. Vinters, and T. J. Montine, "National institute on Aging-Alzheimer's association guidelines for the neuropathologic assessment of alzheimer's disease," Alzheimers. Dement., vol. 8, pp. 1–13, Jan. 2012.

[298] C. R. Jack, Jr, M. S. Albert, D. S. Knopman, G. M. McKhann, R. A. Sperling, M. C. Carrillo, B. Thies, and C. H. Phelps, "Introduction to the recommendations from the national institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease," Alzheimers. Dement., vol. 7, pp. 257–262, May 2011.

[299] C. R. Jack, Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, H. H. Feldman, G. B. Frisoni, H. Hampel, W. J. Jagust, K. A. Johnson, D. S. Knopman, R. C. Petersen, P. Scheltens, R. A. Sperling, and

B. Dubois, "A/T/N: An unbiased descriptive classification scheme for alzheimer disease biomarkers," Neurology, vol. 87, pp. 539–547, Aug. 2016.

[300] K. Strimbu and J. A. Tavel, "What are biomarkers?," Curr. Opin. HIV AIDS, vol. 5, pp. 463–466, Nov. 2010.

[301] J. Cummings, H. H. Feldman, and P. Scheltens, "The "rights" of precision drug development for alzheimer's disease," Alzheimers. Res. Ther., vol. 11, p. 76, Aug. 2019.

[302] B. Dubois, N. Villain, G. B. Frisoni, G. D. Rabinovici, M. Sabbagh, S. Cappa, A. Bejanin, S. Bombois, S. Epelbaum, M. Teichmann, M.-O. Habert, A. Nordberg, K. Blennow, D. Galasko, Y. Stern, C. C. Rowe, S. Salloway, L. S. Schneider, J. L. Cummings, and H. H. Feldman, "Clinical diagnosis of alzheimer's disease: recommendations of the international working group," Lancet Neurol., vol. 20, pp. 484–496, June 2021.

[303] C. E. Teunissen, I. M. W. Verberk, E. H. Thijssen, L. Vermunt, O. Hansson, H. Zetterberg, W. M. van der Flier, M. M. Mielke, and M. Del Campo, "Blood-based biomarkers for alzheimer's disease: towards clinical implementation," Lancet Neurol., vol. 21, pp. 66–77, Jan. 2022.

[304] J. Fang, A. A. Pieper, R. Nussinov, G. Lee, L. Bekris, J. B. Leverenz, J. Cummings, and F. Cheng, "Harnessing endophenotypes and network medicine for alzheimer's drug repurposing," Med. Res. Rev., vol. 40, pp. 2386–2426, Nov. 2020.

[305] S. Budd Haeberlein, P. S. Aisen, F. Barkhof, S. Chalkias, T. Chen, S. Cohen, G. Dent, O. Hansson, K. Harrison, C. von Hehn, T. Iwatsubo, C. Mallinckrodt, C. J. Mummery, K. K. Muralidharan, I. Nestorov, L. Nisenbaum, R. Rajagovindan, L. Skordos, Y. Tian, C. H. van Dyck, B. Vellas, S. Wu, Y. Zhu, and A. Sandrock, "Two randomized phase 3 studies of aducanumab in early alzheimer's disease," J Prev Alzheimers Dis, vol. 9, no. 2, pp. 197–210, 2022.

[306] C. H. van Dyck, C. J. Swanson, P. Aisen, R. J. Bateman, C. Chen, M. Gee, M. Kanekiyo, D. Li, L. Reyderman, S. Cohen, L. Froelich, S. Katayama, M. Sabbagh, B. Vellas, D. Watson, S. Dhadda, M. Irizarry, L. D. Kramer, and T. Iwatsubo, "Lecanemab in early alzheimer's disease," N. Engl. J. Med., vol. 388, pp. 9–21, Jan. 2023.

[307] C. Ballard, D. Aarsland, J. Cummings, J. O'Brien, R. Mills, J. L. Molinuevo, T. Fladby, G. Williams, P. Doherty, A. Corbett, and J. Sultana, "Drug repositioning and repurposing for alzheimer disease," Nat. Rev. Neurol., vol. 16, pp. 661–673, Dec. 2020.

[308] S. Mukherjee, T. M. Perumal, K. Daily, S. K. Sieberts, L. Omberg, C. Preuss, G. W. Carter, L. M. Mangravite, and B. A. Logsdon, "Identifying and ranking potential driver genes of alzheimer's disease using multiview evidence aggregation," Bioinformatics, vol. 35, pp. i568–i576, July 2019.

[309] A. K. Greenwood, J. Gockley, K. Daily, D. Aluthgamage, Z. Leanza, S. K. Sieberts, K. H. Woo, S. Simon, A. Gendel, T. Thyer, K. Do, M. Doerr, M. A. Peters, L. Omberg, B. A. Logsdon, and L. M. Mangravite, "Agora: An open platform for exploration of alzheimer's disease evidence," Alzheimers. Dement., vol. 16, Dec. 2020.

[310] Y. Zhou, J. Fang, L. M. Bekris, Y. H. Kim, A. A. Pieper, J. B. Leverenz, J. Cummings, and F. Cheng, "AlzGPS: a genome-wide positioning systems platform to catalyze multi-omics for alzheimer's drug discovery," Alzheimers. Res. Ther., vol. 13, p. 24, Jan. 2021.

[311] C.-X. Lin, H.-D. Li, C. Deng, S. Erhardt, J. Wang, X. Peng, and J. Wang, "AlzCode: a platform for multiview analysis of genes related to alzheimer's disease," Bioinformatics, Jan. 2022.

[312] D. S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S. L. Chen, D. Hadley, A. Green, P. Khankhanian, and S. E. Baranzini, "Systematic integration of biomedical knowledge prioritizes drugs for repurposing," Elife, vol. 6, pp. 1–35, Sept. 2017.

[313] S. Sadegh, J. Skelton, E. Anastasi, J. Bernett, D. B. Blumenthal, G. Galindez, M. Salgado-Albarrán, O. Lazareva, K. Flanagan, S. Cockell, C. Nogales, A. I. Casas, H. H. H. W. Schmidt, J. Baumbach, A. Wipat, and T. Kacprowski, "Network medicine for disease module identification and drug repurposing with the NeDRex platform," Nat. Commun., vol. 12, p. 6848, Nov. 2021.

[314] D. A. Bennett, J. A. Schneider, Z. Arvanitakis, and R. S. Wilson, "Overview and findings from the religious orders study," Curr. Alzheimer Res., vol. 9, pp. 628–645, July 2012.

[315] D. A. Bennett, J. A. Schneider, A. S. Buchman, L. L. Barnes, P. A. Boyle, and R. S. Wilson, "Overview and findings from the rush memory and aging project," Curr. Alzheimer Res., vol. 9, pp. 646–663, July 2012.

[316] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The alzheimer's disease neuroimaging initiative," Neuroimaging Clin. N. Am., vol. 15, pp. 869–77, xi–xii, Nov. 2005.

[317] M. W. Weiner, P. S. Aisen, C. R. Jack, Jr, W. J. Jagust, J. Q. Trojanowski, L. Shaw, A. J. Saykin, J. C. Morris, N. Cairns, L. A. Beckett, A. Toga, R. Green, S. Walter, H. Soares, P. Snyder, E. Siemers, W. Potter, P. E. Cole, M. Schmidt, and Alzheimer's Disease Neuroimaging Initiative, "The alzheimer's disease neuroimaging initiative: progress report and future plans," Alzheimers. Dement., vol. 6, pp. 202–11.e7, May 2010.

[318] K. L. Crawford, S. C. Neu, and A. W. Toga, "The image and data archive at the laboratory of neuro imaging," Neuroimage, vol. 124, pp. 1080–1083, Jan. 2016.

[319] N. W. Shock, Normal human aging: The Baltimore longitudinal study of aging. US Department of Health and Human Services, Public Health Service, National . . . , 1984.

[320] L. Ferrucci, "The baltimore longitudinal study of aging (BLSA): a 50-year-long journey and plans for the future," J. Gerontol. A Biol. Sci. Med. Sci., vol. 63, pp. 1416–1419, Dec. 2008.

[321] R. J. O'Brien, S. M. Resnick, A. B. Zonderman, L. Ferrucci, B. J. Crain, O. Pletnikova, G. Rudow, D. Iacono, M. A. Riudavets, I. Driscoll, D. L. Price, L. J. Martin, and J. C. Troncoso, "Neuropathologic studies of the baltimore longitudinal study of aging (BLSA)," 2009.

[322] M. Wang, N. D. Beckmann, P. Roussos, E. Wang, X. Zhou, Q. Wang, C. Ming, R. Neff, W. Ma, J. F. Fullard, M. E. Hauberg, J. Bendl, M. A. Peters, B. Logsdon, P. Wang, M. Mahajan, L. M. Mangravite, E. B. Dammer, D. M. Duong, J. J. Lah, N. T. Seyfried, A. I. Levey, J. D. Buxbaum, M. Ehrlich, S. Gandy, P. Katsel, V. Haroutunian, E. Schadt, and B. Zhang, "The mount sinai cohort of large-scale genomic, transcriptomic and proteomic data in alzheimer's disease," Sci Data, vol. 5, p. 180185, Sept. 2018.

[323] T. G. Beach, C. H. Adler, L. I. Sue, G. Serrano, H. A. Shill, D. G. Walker, L. Lue, A. E. Roher, B. N. Dugger, C. Maarouf, A. C. Birdsill, A. Intorcia, M. Saxon-Labelle, J. Pullen, A. Scroggins, J. Filon, S. Scott, B. Hoffman, A. Garcia, J. N. Caviness, J. G. Hentz, E. Driver-Dunckley, S. A. Jacobson, K. J. Davis, C. M. Belden, K. E. Long, M. Malek-Ahmadi, J. J. Powell, L. D. Gale, L. R. Nicholson, R. J. Caselli, B. K. Woodruff, S. Z. Rapscak, G. L. Ahern, J. Shi, A. D. Burke, E. M. Reiman, and M. N. Sabbagh, "Arizona study of aging and neurodegenerative disorders and brain and body donation program," Neuropathology, vol. 35, pp. 354–389, Aug. 2015.

[324] W. A. Kukull, R. Higdon, J. D. Bowen, W. C. McCormick, L. Teri, G. D. Schellenberg, G. van Belle, L. Jolley, and E. B. Larson, "Dementia and alzheimer disease incidence: a prospective cohort study," Arch. Neurol., vol. 59, pp. 1737–1746, Nov. 2002.

[325] A. K. Greenwood, K. S. Montgomery, N. Kauer, K. H. Woo, Z. J. Leanza, W. L. Poehlman, J. Gockley, S. K. Sieberts, L. Bradic, B. A. Logsdon, M. A. Peters, L. Omberg, and L. M. Mangravite, "The AD knowledge portal: A repository for Multi-Omic data on alzheimer's disease and aging," Curr. Protoc. Hum. Genet., vol. 108, p. e105, Dec. 2020.

[326] S. H. Friend and T. C. Norman, "Metcalfe's law and the biology information commons," Nat. Biotechnol., vol. 31, pp. 297–303, Apr. 2013.

[327] S. K. Sieberts, T. M. Perumal, M. M. Carrasquillo, M. Allen, J. S. Reddy, G. E. Hoffman, K. K. Dang, J. Calley, P. J. Ebert, J. Eddy, X. Wang, A. K. Greenwood, S. Mostafavi, CommonMind Consortium (CMC), The AMP-AD Consortium, L. Omberg, M. A. Peters, B. A. Logsdon, P. L. De Jager, N. Ertekin-Taner, and L. M. Mangravite, "Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions," Sci Data, vol. 7, p. 340, Oct. 2020.

[328] Y.-W. Wan, R. Al-Ouran, C. G. Mangleburg, T. M. Perumal, T. V. Lee, K. Allison, V. Swarup, C. C. Funk, C. Gaiteri, M. Allen, M. Wang, S. M. Neuner, C. C. Kaczorowski, V. M. Philip, G. R. Howell, H. Martini-Stoica, H. Zheng, H. Mei, X. Zhong, J. W. Kim, V. L. Dawson, T. M. Dawson, P.-C. Pao, L.-H. Tsai, J.-V. Haure-Mirande, M. E. Ehrlich, P. Chakrabarty, Y. Levites, X. Wang, E. B. Dammer, G. Srivastava, S. Mukherjee, S. K. Sieberts, L. Omberg, K. D. Dang, J. A. Eddy, P. Snyder, Y. Chae, S. Amberkar, W. Wei, W. Hide, C. Preuss, A. Ergun, P. J. Ebert, D. C. Airey, S. Mostafavi, L. Yu, H.-U. Klein, Accelerating Medicines Partnership-Alzheimer's Disease Consortium, G. W. Carter, D. A. Collier, T. E. Golde, A. I. Levey, D. A. Bennett, K. Estrada, T. M. Townsend, B. Zhang, E. Schadt, P. L. De Jager, N. D. Price, N. Ertekin-Taner, Z. Liu, J. M. Shulman, L. M. Mangravite, and B. A. Logsdon, "Meta-Analysis of the alzheimer's disease human brain transcriptome and functional dissection in mouse models," Cell Rep., vol. 32, p. 107908, July 2020.

[329] E. C. B. Johnson, E. B. Dammer, D. M. Duong, L. Ping, M. Zhou, L. Yin, L. A. Higginbotham, A. Guajardo, B. White, J. C. Troncoso, M. Thambisetty, T. J. Montine, E. B. Lee, J. Q. Trojanowski, T. G. Beach, E. M. Reiman, V. Haroutunian, M. Wang, E. Schadt, B. Zhang, D. W. Dickson, N. Ertekin-Taner, T. E. Golde, V. A. Petyuk, P. L. De Jager, D. A. Bennett, T. S. Wingo, S. Rangaraju, I. Hajjar, J. M. Shulman, J. J. Lah, A. I. Levey, and N. T. Seyfried, "Large-scale proteomic analysis of alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation," Nat. Med., vol. 26, pp. 769–780, May 2020.

[330] E. C. B. Johnson, E. K. Carter, E. B. Dammer, D. M. Duong, E. S. Gerasimov, Y. Liu, J. Liu, R. Betarbet, L. Ping, L. Yin, G. E. Serrano, T. G. Beach, J. Peng, P. L. De Jager, V. Haroutunian, B. Zhang, C. Gaiteri, D. A. Bennett, M. Gearing, T. S. Wingo, A. P. Wingo, J. J. Lah, A. I. Levey, and N. T. Seyfried, "Large-scale deep multi-layer analysis of alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level," Nat. Neurosci., vol. 25, pp. 213–225, Feb. 2022.

[331] A. Kuzma, O. Valladares, R. Cweibel, E. Greenfest-Allen, D. M. Childress, J. Malamon, P. Gangadharan, Y. Zhao, L. Qu, Y. Y. Leung, A. C. Naj, C. J. Stoeckert, Jr, G. D. Schellenberg, and L.-S. Wang, "NIAGADS: The NIA genetics of alzheimer's disease data storage site," Alzheimers. Dement., vol. 12, pp. 1200–1203, Nov. 2016.

[332] F. Cunningham, P. Achuthan, W. Akanni, J. Allen, M. R. Amode, I. M. Armean, R. Bennett, J. Bhai, K. Billis, S. Boddu, C. Cummins, C. Davidson, K. J. Dodiya, A. Gall, C. G. Girón, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, J. C. Marugán, T. Maurel, A. C. McMahon, B. Moore, J. Morales, J. M. Mudge, M. Nuhn, D. Ogeh, A. Parker, A. Parton, M. Patricio, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, H. Sparrow, E. Stapleton, M. Szuba, K. Taylor, G. Threadgold, A. Thormann, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, A. D. Yates, D. R. Zerbino, and P. Flicek, "Ensembl 2019," Nucleic Acids Res., vol. 47, pp. D745–D751, Jan. 2019.

[333] E. Birney, T. D. Andrews, P. Bevan, M. Caccamo, Y. Chen, L. Clarke, G. Coates, J. Cuff, V. Curwen, T. Cutts, T. Down, E. Eyras, X. M. Fernandez-Suarez, P. Gane, B. Gibbins, J. Gilbert, M. Hammond, H.-R. Hotz, V. Iyer, K. Jekosch, A. Kahari, A. Kasprzyk, D. Keefe, S. Keenan, H. Lehvaslaiho, G. McVicker, C. Melsopp, P. Meidl, E. Mongin, R. Pettett, S. Potter, G. Proctor, M. Rae, S. Searle, G. Slater, D. Smedley, J. Smith, W. Spooner, A. Stabenau, J. Stalker, R. Storey, A. Ureta-Vidal, K. C. Woodwark, G. Cameron, R. Durbin, A. Cox, T. Hubbard, and M. Clamp, "An overview of ensembl," Genome Res., vol. 14, pp. 925–928, May 2004.

[334] B. Braschi, P. Denny, K. Gray, T. Jones, R. Seal, S. Tweedie, B. Yates, and E. Bruford, "Genenames.org: the HGNC and VGNC resources in 2019," Nucleic Acids Res., vol. 47, pp. D786–D792, Jan. 2019.

[335] R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B.-K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos, "The accessible chromatin landscape of the human genome," Nature, vol. 489, pp. 75–82, Sept. 2012.

[336] FANTOM Consortium and the RIKEN PMI and CLST (DGT), A. R. R. Forrest, H. Kawaji, M. Rehli, J. K. Baillie, M. J. L. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jørgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. A. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. A. C. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. M. Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drabløs, A. S. B. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furino, J.-I. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. Ho Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. P. Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. J. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-Sim, R.-I. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, S. Noma, T. Noazaki, S. Ogishima, N. Ohkura, H. Ohimiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. D. Prendergast, O. J. L. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. C. 't Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tat-

sukawa, Z. Tatum, M. Thompson, H. Toyodo, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verado, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci, and Y. Hayashizaki, "A promoter-level mammalian expression atlas," Nature, vol. 507, pp. 462–470, Mar. 2014.

[337] T. Long, M. Hicks, H.-C. Yu, W. H. Biggs, E. F. Kirkness, C. Menni, J. Zierer, K. S. Small, M. Mangino, H. Messier, S. Brewerton, Y. Turpaz, B. A. Perkins, A. M. Evans, L. A. D. Miller, L. Guo, C. T. Caskey, N. J. Schork, C. Garner, T. D. Spector, J. C. Venter, and A. Telenti, "Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites," Nat. Genet., vol. 49, pp. 568–578, Apr. 2017.

[338] G. W. Beecham, K. Hamilton, A. C. Naj, E. R. Martin, M. Huentelman, A. J. Myers, J. J. Corneveaux, J. Hardy, J.-P. Vonsattel, S. G. Younkin, D. A. Bennett, P. L. De Jager, E. B. Larson, P. K. Crane, M. I. Kamboh, J. K. Kofler, D. C. Mash, L. Duque, J. R. Gilbert, H. Gwirtsman, J. D. Buxbaum, P. Kramer, D. W. Dickson, L. A. Farrer, M. P. Frosch, B. Ghetti, J. L. Haines, B. T. Hyman, W. A. Kukull, R. P. Mayeux, M. A. Pericak-Vance, J. A. Schneider, J. Q. Trojanowski, E. M. Reiman, Alzheimer's Disease Genetics Consortium (ADGC), G. D. Schellenberg, and T. J. Montine, "Genome-wide association meta-analysis of neuropathologic features of alzheimer's disease and related dementias," PLoS Genet., vol. 10, p. e1004606, Sept. 2014.

[339] Y. Deming, J. Xia, Y. Cai, J. Lord, P. Holmans, S. Bertelsen, D. Holtzman, J. C. Morris, K. Bales, E. H. Pickering, J. Kauwe, A. Goate, C. Cruchaga, and Alzheimer's Disease Neuroimaging Initiative (ADNI), "A potential endophenotype for alzheimer's disease: cerebrospinal fluid clusterin," Neurobiol. Aging, vol. 37, pp. 208.e1–208.e9, Jan. 2016.

[340] Y. Deming, Z. Li, M. Kapoor, O. Harari, J. L. Del-Aguila, K. Black, D. Carrell, Y. Cai, M. V. Fernandez, J. Budde, S. Ma, B. Saef, B. Howells, K.-L. Huang, S. Bertelsen, A. M. Fagan, D. M. Holtzman, J. C. Morris, S. Kim, A. J. Saykin, P. L. De Jager, M. Albert, A. Moghekar, R. O'Brien, M. Riemenschneider, R. C. Petersen, K. Blennow, H. Zetterberg, L. Minthon, V. M. Van Deerlin, V. M.-Y. Lee, L. M. Shaw, J. Q. Trojanowski, G. Schellenberg, J. L. Haines, R. Mayeux, M. A. Pericak-Vance, L. A. Farrer, E. R. Peskind, G. Li, A. F. Di Narzo, Alzheimer's Disease Neuroimaging Initiative (ADNI), Alzheimer Disease Genetic Consortium (ADGC), J. S. K. Kauwe, A. M. Goate, and C. Cruchaga, "Genome-wide association study identifies four novel loci associated with alzheimer's endophenotypes and disease modifiers," Acta Neuropathol., vol. 133, pp. 839–856, May 2017.

[341] K.-L. Huang, E. Marcora, A. A. Pimenova, A. F. Di Narzo, M. Kapoor, S. C. Jin, O. Harari, S. Bertelsen, B. P. Fairfax, J. Czajkowski, V. Chouraki, B. Grenier-Boley, C. Bellenguez, Y. Deming, A. McKenzie, T. Raj, A. E. Renton, J. Budde, A. Smith, A. Fitzpatrick, J. C. Bis, A. DeStefano, H. H. H. Adams, M. A. Ikram, S. van der Lee, J. L. Del-Aguila, M. V. Fernandez, L. Ibañez, International Genomics of Alzheimer's Project, Alzheimer's Disease Neuroimaging Initiative, R. Sims, V. Escott-Price, R. Mayeux, J. L. Haines, L. A. Farrer, M. A. Pericak-Vance, J. C. Lambert, C. van Duijn, L. Launer, S. Seshadri, J. Williams, P. Amouyel, G. D. Schellenberg, B. Zhang, I. Borecki, J. S. K. Kauwe, C. Cruchaga, K. Hao, and A. M. Goate, "A common haplotype lowers PU.1 expression in myeloid cells and delays onset of alzheimer's disease," Nat. Neurosci., vol. 20, pp. 1052–1061, Aug. 2017.

[342] D. P. Wightman, I. E. Jansen, J. E. Savage, A. A. Shadrin, S. Bahrami, D. Holland, A. Rongve, S. Børte, B. S. Winsvold, O. K. Drange, A. E. Martinsen, A. H. Skogholt, C. Willer, G. Bråthen, I. Bosnes, J. B. Nielsen, L. G. Fritsche, L. F. Thomas, L. M. Pedersen, M. E. Gabrielsen, M. B. Johnsen, T. W. Meisingset, W. Zhou, P. Proitsi, A. Hodges, R. Dobson, L. Velayudhan, K. Heilbron,

A. Auton, 23andMe Research Team, J. M. Sealock, L. K. Davis, N. L. Pedersen, C. A. Reynolds, I. K. Karlsson, S. Magnusson, H. Stefansson, S. Thordardottir, P. V. Jonsson, J. Snaedal, A. Zettergren, I. Skoog, S. Kern, M. Waern, H. Zetterberg, K. Blennow, E. Stordal, K. Hveem, J.-A. Zwart, L. Athanasiu, P. Selnes, I. Saltvedt, S. B. Sando, I. Ulstein, S. Djurovic, T. Fladby, D. Aarsland, G. Selbæk, S. Ripke, K. Stefansson, O. A. Andreassen, and D. Posthuma, "A genome-wide association study with 1,126,563 individuals identifies new risk loci for alzheimer's disease," Nat. Genet., vol. 53, pp. 1276–1282, Sept. 2021.

[343] P. L. De Jager, Y. Ma, C. McCabe, J. Xu, B. N. Vardarajan, D. Felsky, H.-U. Klein, C. C. White, M. A. Peters, B. Lodgson, P. Nejad, A. Tang, L. M. Mangravite, L. Yu, C. Gaiteri, S. Mostafavi, J. A. Schneider, and D. A. Bennett, "A multi-omic atlas of the human frontal cortex for aging and alzheimer's disease research," Sci Data, vol. 5, p. 180142, Aug. 2018.

[344] M. Allen, M. M. Carrasquillo, C. Funk, B. D. Heavner, F. Zou, C. S. Younkin, J. D. Burgess, H.-S. Chai, J. Crook, J. A. Eddy, H. Li, B. Logsdon, M. A. Peters, K. K. Dang, X. Wang, D. Serie, C. Wang, T. Nguyen, S. Lincoln, K. Malphrus, G. Bisceglio, M. Li, T. E. Golde, L. M. Mangravite, Y. Asmann, N. D. Price, R. C. Petersen, N. R. Graff-Radford, D. W. Dickson, S. G. Younkin, and N. Ertekin-Taner, "Human whole genome genotype and transcriptome data for alzheimer's and other neurodegenerative diseases," Sci Data, vol. 3, p. 160089, Oct. 2016.

[345] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, J. C. Morris, R. C. Petersen, J. Salazar, A. J. Saykin, L. M. Shaw, A. W. Toga, and J. Q. Trojanowski, "The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement," 2017.

[346] I. Robinson, J. Webber, J. Webber, and E. Eifrem, Graph Databases. O'Reilly, 2013.

[347] S. T. Sherry, M. Ward, and K. Sirotkin, "dbSNP—Database for single nucleotide polymorphisms and other classes of minor genetic variation," Genome Res., vol. 9, pp. 677–679, Aug. 1999.

[348] UniProt Consortium, "UniProt: a worldwide hub of protein knowledge," Nucleic Acids Res., vol. 47, pp. D506–D515, Jan. 2019.

[349] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: online learning of social representations," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '14, (New York, NY, USA), pp. 701–710, Association for Computing Machinery, Aug. 2014.

[350] R. Ietswaart, B. M. Gyori, J. A. Bachman, P. K. Sorger, and L. S. Churchman, "GeneWalk identifies relevant gene functions for a biological context using network representation learning," Genome Biol., vol. 22, p. 55, Feb. 2021.

[351] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," arXiv [stat.ML], Feb. 2018.

[352] W. Gu, A. Tandon, Y.-Y. Ahn, and F. Radicchi, "Principled approach to the selection of the embedding dimension of networks," Nat. Commun., vol. 12, p. 3772, June 2021.

[353] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," KDD, vol. 2016, pp. 855–864, Aug. 2016.

[354] Z. Xie, A. Bailey, M. V. Kuleshov, D. J. B. Clarke, J. E. Evangelista, S. L. Jenkins, A. Lachmann, M. L. Wojciechowicz, E. Kropiwnicki, K. M. Jagodnik, M. Jeon, and A. Ma'ayan, "Gene set knowledge discovery with enrichr," Curr Protoc, vol. 1, p. e90, Mar. 2021.

[355] R. Edgar, M. Domrachev, and A. E. Lash, "Gene expression omnibus: NCBI gene expression and hybridization array data repository," Nucleic Acids Res., vol. 30, pp. 207–210, Jan. 2002.

[356] M. Gillespie, B. Jassal, R. Stephan, M. Milacic, K. Rothfels, A. Senff-Ribeiro, J. Griss, C. Sevilla, L. Matthews, C. Gong, C. Deng, T. Varusai, E. Ragueneau, Y. Haider, B. May, V. Shamovsky, J. Weiser, T. Brunson, N. Sanati, L. Beckman, X. Shao, A. Fabregat, K. Sidiropoulos, J. Murillo, G. Viteri, J. Cook, S. Shorser, G. Bader, E. Demir, C. Sander, R. Haw, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio, "The reactome pathway knowledgebase 2022," Nucleic Acids Res., vol. 50, pp. D687–D692, Jan. 2022.

[357] M. Martens, A. Ammar, A. Riutta, A. Waagmeester, D. N. Slenter, K. Hanspers, R. A Miller, D. Digles, E. N. Lopes, F. Ehrhart, L. J. Dupuis, L. A. Winckers, S. L. Coort, E. L. Willighagen, C. T. Evelo, A. R. Pico, and M. Kutmon, "WikiPathways: connecting communities," Nucleic Acids Res., vol. 49, pp. D613–D621, Jan. 2021.

[358] L. Kolberg, U. Raudvere, I. Kuzmin, J. Vilo, and H. Peterson, "gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:profiler," F1000Res., vol. 9, p. 709, July 2020.

[359] A. Alexa and J. Rahnenfuhrer, "topGO: Enrichment analysis for gene ontology," 2019.

[360] T. A. Knijnenburg, L. F. A. Wessels, M. J. T. Reinders, and I. Shmulevich, "Fewer permutations, more accurate p-values," Bioinformatics, vol. 25, pp. i161–8, June 2009.

[361] J. M. Long and D. M. Holtzman, "Alzheimer disease: An update on pathobiology and treatment strategies," Cell, vol. 179, pp. 312–339, Oct. 2019.

[362] A. Mullard, "Landmark alzheimer's drug approval confounds research community," Nature, vol. 594, pp. 309–310, June 2021.

[363] D. Ferreira, L.-O. Wahlund, and E. Westman, "The heterogeneity within alzheimer's disease," Aging, vol. 10, pp. 3058–3060, Nov. 2018.

[364] G. M. Sancesario and S. Bernardini, "Alzheimer's disease in the omics era," Clin. Biochem., vol. 59, pp. 9–16, Sept. 2018.

[365] B. Lee, S. Zhang, A. Poleksic, and L. Xie, "Heterogeneous Multi-Layered network model for omics data integration and analysis," Front. Genet., vol. 10, no. January, p. 1381, 2019.

[366] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," IEEE Data Eng. Bull., vol. 40, pp. 52–74, Sept. 2017.

[367] W. Nelson, M. Zitnik, B. Wang, J. Leskovec, A. Goldenberg, and R. Sharan, "To embed or not: Network embedding as a paradigm in computational biology," Front. Genet., vol. 10, p. 381, May 2019.

[368] H. Cai, V. W. Zheng, and K. C.-C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," IEEE Trans. Knowl. Data Eng., vol. 30, pp. 1616–1637, Sept. 2018.

[369] Y. Shi and D. M. Holtzman, "Interplay between innate immunity and alzheimer disease: APOE and TREM2 in the spotlight," Nat. Rev. Immunol., vol. 18, pp. 759–772, Dec. 2018.

[370] J.-C. Lambert, S. Heath, G. Even, D. Campion, K. Sleegers, M. Hiltunen, O. Combarros, D. Zelenika, M. J. Bullido, B. Tavernier, L. Letenneur, K. Bettens, C. Berr, F. Pasquier, N. Fiévet, P. Barberger-Gateau, S. Engelborghs, P. De Deyn, I. Mateo, A. Franck, S. Helisalmi, E. Porcellini, O. Hanon, European Alzheimer's Disease Initiative Investigators, M. M. de Pancorbo, C. Lendon, C. Dufouil, C. Jaillard, T. Leveillard, V. Alvarez, P. Bosco, M. Mancuso, F. Panza, B. Nacmias, P. Bossù, P. Piccardi, G. Annoni, D. Seripa, D. Galimberti, D. Hannequin, F. Licastro, H. Soininen, K. Ritchie, H. Blanché, J.-F. Dartigues, C. Tzourio, I. Gut, C. Van Broeckhoven, A. Alpérovitch, M. Lathrop, and P. Amouyel, "Genome-wide association study identifies variants at CLU and CR1 associated with alzheimer's disease," Nat. Genet., vol. 41, pp. 1094–1099, Oct. 2009.

[371] D. Harold, R. Abraham, P. Hollingworth, R. Sims, A. Gerrish, M. L. Hamshere, J. S. Pahwa, V. Moskvina, K. Dowzell, A. Williams, N. Jones, C. Thomas, A. Stretton, A. R. Morgan, S. Lovestone, J. Powell, P. Proitsi, M. K. Lupton, C. Brayne, D. C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. Morgan, K. S. Brown, P. A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A. D. Smith, S. Love, P. G. Kehoe, J. Hardy, S. Mead, N. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, B. Schürmann, R. Heun, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M. Dichgans, L. Frölich, H. Hampel, M. Hüll, D. Rujescu, A. M. Goate, J. S. K. Kauwe, C. Cruchaga, P. Nowotny, J. C. Morris, K. Mayo, K. Sleegers, K. Bettens, S. Engelborghs, P. P. De Deyn, C. Van Broeckhoven, G. Livingston, N. J. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C. E. Shaw, M. Tsolaki, A. B. Singleton, R. Guerreiro, T. W. Mühleisen, M. M. Nöthen, S. Moebus, K.-H. Jöckel, N. Klopp, H.-E. Wichmann, M. M. Carrasquillo, V. S. Pankratz, S. G. Younkin, P. A. Holmans, M. O'Donovan, M. J. Owen, and J. Williams, "Genome-wide association study identifies variants at CLU and PICALM associated with alzheimer's disease," Nat. Genet., vol. 41, pp. 1088–1093, Oct. 2009.

[372] M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, and A. Ma'ayan, "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update," Nucleic Acids Res., vol. 44, pp. W90–7, July 2016.

[373] J. Bauzon, G. Lee, and J. Cummings, "Repurposed agents in the alzheimer's disease drug development pipeline," Alzheimers. Res. Ther., vol. 12, p. 98, Aug. 2020.

[374] M. Gold, C. Alderton, M. Zvartau-Hind, S. Egginton, A. M. Saunders, M. Irizarry, S. Craft, G. Landreth, U. Linnamägi, and S. Sawchak, "Rosiglitazone monotherapy in mild-to-moderate alzheimer's disease: results from a randomized, double-blind, placebo-controlled phase III study," Dement. Geriatr. Cogn. Disord., vol. 30, pp. 131–146, Aug. 2010.

[375] C. Harrington, S. Sawchak, C. Chiang, J. Davies, C. Donovan, A. M. Saunders, M. Irizarry, B. Jeter, M. Zvartau-Hind, C. H. van Dyck, and M. Gold, "Rosiglitazone does not improve cognition or global function when used as adjunctive therapy to AChE inhibitors in mild-to-moderate alzheimer's disease: two phase 3 studies," Curr. Alzheimer Res., vol. 8, pp. 592–606, Aug. 2011.

[376] P. N. Tariot, L. S. Schneider, J. Cummings, R. G. Thomas, R. Raman, L. J. Jakimovich, R. Loy, B. Bartocci, A. Fleisher, M. S. Ismail, A. Porsteinsson, M. Weiner, C. R. Jack, Jr, L. Thal, P. S. Aisen, and Alzheimer's Disease Cooperative Study Group, "Chronic divalproex sodium to attenuate agitation and clinical progression of alzheimer disease," Arch. Gen. Psychiatry, vol. 68, pp. 853–861, Aug. 2011.

[377] B. Mdawar, E. Ghossoub, and R. Khoury, "Selective serotonin reuptake inhibitors and alzheimer's disease," Neural Regeneration Res., vol. 15, pp. 41–46, Jan. 2020.

[378] Y. Xie, P.-P. Liu, Y.-J. Lian, H.-B. Liu, and J.-S. Kang, "The effect of selective serotonin reuptake inhibitors on cognitive function in patients with alzheimer's disease and vascular dementia: focusing on fluoxetine with long follow-up periods," Signal Transduct Target Ther, vol. 4, p. 30, Aug. 2019.

[379] J. Cummings, G. Lee, A. Ritter, M. Sabbagh, and K. Zhong, "Alzheimer's disease drug development pipeline: 2020," Alzheimers. Dement., vol. 6, p. e12050, July 2020.

[380] I. Hajjar, M. Okafor, D. McDaniel, M. Obideen, E. Dee, M. Shokouhi, A. A. Quyyumi, A. Levey, and F. Goldstein, "Effects of candesartan vs lisinopril on neurocognitive function in older adults with executive mild cognitive impairment: A randomized clinical trial," JAMA Netw Open, vol. 3, p. e2012252, Aug. 2020.

[381] I. Hajjar, M. Hart, Y.-L. Chen, W. Mack, W. Milberg, H. Chui, and L. Lipsitz, "Effect of antihypertensive therapy on cognitive function in early executive cognitive impairment: a double-blind randomized clinical trial," Arch. Intern. Med., vol. 172, pp. 442–444, Mar. 2012.

[382] C. S. Rosenfeld, D. A. Shay, and V. J. Vieira-Potter, "Cognitive effects of aromatase and possible role in memory disorders," Front. Endocrinol., vol. 9, p. 610, Oct. 2018.

[383] B. Decourt, D. K. Lahiri, and M. N. Sabbagh, "Targeting tumor necrosis factor alpha for alzheimer's disease," Curr. Alzheimer Res., vol. 14, no. 4, pp. 412–425, 2017.

[384] J. Butchart, L. Brook, V. Hopkins, J. Teeling, U. Püntener, D. Culliford, R. Sharples, S. Sharif, B. McFarlane, R. Raybould, R. Thomas, P. Passmore, V. H. Perry, and C. Holmes, "Etanercept in alzheimer disease: A randomized, placebo-controlled, double-blind, phase 2 trial," Neurology, vol. 84, pp. 2161–2168, May 2015.

[385] N. Torres-Acosta, J. H. O'Keefe, E. L. O'Keefe, R. Isaacson, and G. Small, "Therapeutic potential of TNF-α inhibition for alzheimer's disease prevention," J. Alzheimers. Dis., vol. 78, no. 2, pp. 619–626, 2020.

[386] M. E. De Sousa Rodrigues, M. C. Houser, D. I. Walker, D. P. Jones, J. Chang, C. J. Barnum, and M. G. Tansey, "Targeting soluble tumor necrosis factor as a potential intervention to lower risk for late-onset alzheimer's disease associated with obesity, metabolic syndrome, and type 2 diabetes," Alzheimers. Res. Ther., vol. 12, p. 1, Dec. 2019.

[387] K. P. MacPherson, P. Sompol, G. T. Kannarkat, J. Chang, L. Sniffen, M. E. Wildner, C. M. Norris, and M. G. Tansey, "Peripheral administration of the soluble TNF inhibitor XPro1595 modifies brain immune cell profiles, decreases beta-amyloid plaque load, and rescues impaired long-term potentiation in 5xFAD mice," Neurobiol. Dis., vol. 102, pp. 81–95, June 2017.

[388] N. Z. R. Steele, J. S. Carr, L. W. Bonham, E. G. Geier, V. Damotte, Z. A. Miller, R. S. Desikan, K. L. Boehme, S. Mukherjee, P. K. Crane, J. S. K. Kauwe, J. H. Kramer, B. L. Miller, G. Coppola, J. A. Hollenbach, Y. Huang, and J. S. Yokoyama, "Fine-mapping of the human leukocyte antigen locus as a risk factor for alzheimer disease: A case-control study," PLoS Med., vol. 14, p. e1002272, Mar. 2017.

[389] A. O. Sodero and F. J. Barrantes, "Pleiotropic effects of statins on brain cells," Biochim. Biophys. Acta Biomembr., vol. 1862, p. 183340, Sept. 2020.

[390] J. K. Liao and U. Laufs, "Pleiotropic effects of statins," Annu. Rev. Pharmacol. Toxicol., vol. 45, no. 1, pp. 89–118, 2005.

[391] M. D. M. Haag, A. Hofman, P. J. Koudstaal, B. H. C. Stricker, and M. M. B. Breteler, "Statins are associated with a reduced risk of alzheimer disease regardless of lipophilicity. the rotterdam study," J. Neurol. Neurosurg. Psychiatry, vol. 80, pp. 13–17, Jan. 2009.

[392] G. Torrandell-Haro, G. L. Branigan, F. Vitali, N. Geifman, J. M. Zissimopoulos, and R. D. Brinton, "Statin therapy and risk of alzheimer's and age-related neurodegenerative diseases," Alzheimers. Dement., vol. 6, p. e12108, Nov. 2020.

[393] C.-S. Chu, P.-T. Tseng, B. Stubbs, T.-Y. Chen, C.-H. Tang, D.-J. Li, W.-C. Yang, Y.-W. Chen, C.-K. Wu, N. Veronese, A. F. Carvalho, B. S. Fernandes, N. Herrmann, and P.-Y. Lin, "Use of statins and the risk of dementia and mild cognitive impairment: A systematic review and meta-analysis," Sci. Rep., vol. 8, p. 5804, Apr. 2018.

[394] J. M. Zissimopoulos, D. Barthold, R. D. Brinton, and G. Joyce, "Sex and race differences in the association between statin use and the incidence of alzheimer disease," JAMA Neurol., vol. 74, pp. 225–232, Feb. 2017.

[395] N. Geifman, R. D. Brinton, R. E. Kennedy, L. S. Schneider, and A. J. Butte, "Evidence for benefit of statins to modify cognitive decline and risk in alzheimer's disease," Alzheimers. Res. Ther., vol. 9, p. 10, Feb. 2017.

[396] B. G. Schultz, D. K. Patten, and D. J. Berlau, "The role of statins in both cognitive impairment and protection against dementia: a tale of two mechanisms," Transl. Neurodegener., vol. 7, p. 5, Feb. 2018.

[397] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, and M. Wilson, "DrugBank 5.0: a major update to the DrugBank database for 2018," Nucleic Acids Res., vol. 46, pp. D1074–D1082, Jan. 2018.

[398] B. L. Walling and M. Kim, "LFA-1 in T cell migration and differentiation," Front. Immunol., vol. 9, p. 952, May 2018.

[399] G. Weitz-Schmidt, K. Welzenbach, V. Brinkmann, T. Kamata, J. Kallen, C. Bruns, S. Cottens, Y. Takada, and U. Hommel, "Statins selectively inhibit leukocyte function antigen-1 by binding to a novel regulatory integrin site," Nat. Med., vol. 7, pp. 687–692, June 2001.

[400] E. Zenaro, E. Pietronigro, V. Della Bianca, G. Piacentino, L. Marongiu, S. Budui, E. Turano, B. Rossi, S. Angiari, S. Dusi, A. Montresor, T. Carlucci, S. Nanì, G. Tosadori, L. Calciano, D. Catalucci, G. Berton, B. Bonetti, and G. Constantin, "Neutrophils promote alzheimer's disease-like pathology and cognitive decline via LFA-1 integrin," Nat. Med., vol. 21, pp. 880–886, Aug. 2015.

[401] J.-V. Haure-Mirande, M. Audrain, T. Fanutza, S. H. Kim, W. L. Klein, C. Glabe, B. Readhead, J. T. Dudley, R. D. Blitzer, M. Wang, B. Zhang, E. E. Schadt, S. Gandy, and M. E. Ehrlich, "Deficiency of TYROBP, an adapter protein for TREM2 and CR3 receptors, is neuroprotective in a mouse model of early alzheimer's pathology," Acta Neuropathol., vol. 134, pp. 769–788, Nov. 2017.

[402] S. Vasiliou, "Oral fingolimod for the treatment of relapsing-remitting multiple sclerosis," Drugs Today, vol. 46, pp. 315–325, May 2010.

[403] I. Carreras, N. Aytan, J.-K. Choi, C. M. Tognoni, N. W. Kowall, B. G. Jenkins, and A. Dedeoglu, "Dual dose-dependent effects of fingolimod in a mouse model of alzheimer's disease," Sci. Rep., vol. 9, p. 10972, July 2019.

[404] H. Jeśko, P. L. Wencel, W. J. Lukiw, and R. P. Strosznajder, "Modulatory effects of fingolimod (FTY720) on the expression of sphingolipid Metabolism-Related genes in an animal model of alzheimer's disease," Mol. Neurobiol., vol. 56, pp. 174–185, Jan. 2019.

[405] S. Ozakbas, B. Piri Cinar, P. Yigit, C. Baba, O. Sagici, and Multiple Sclerosis Research Group, "Five-year real-world data on fingolimod treatment's effects on cognitive function," Mult. Scler. Relat. Disord., vol. 54, p. 103089, Sept. 2021.

[406] S. Mukherjee, L. Heath, C. Preuss, S. Jayadev, G. A. Garden, A. K. Greenwood, S. K. Sieberts, P. L. De Jager, N. Ertekin-Taner, G. W. Carter, L. M. Mangravite, and B. A. Logsdon, "Molecular estimation of neurodegeneration pseudotime in older brains," Nat. Commun., vol. 11, p. 5781, Nov. 2020.

[407] M. L. Bennett, F. C. Bennett, S. A. Liddelow, B. Ajami, J. L. Zamanian, N. B. Fernhoff, S. B. Mulinyawe, C. J. Bohlen, A. Adil, A. Tucker, I. L. Weissman, E. F. Chang, G. Li, G. A. Grant, M. G. Hayden Gephart, and B. A. Barres, "New tools for studying microglia in the mouse and human CNS," Proc. Natl. Acad. Sci. U. S. A., vol. 113, pp. E1738–46, Mar. 2016.

[408] T. K. Ulland and M. Colonna, "TREM2 - a key player in microglial biology and alzheimer disease," Nat. Rev. Neurol., vol. 14, pp. 667–675, Nov. 2018.

[409] G. R. Dohle, M. Smit, and R. F. A. Weber, "Androgens and male fertility," World J. Urol., vol. 21, pp. 341–345, Nov. 2003.

[410] S. R. Davis and S. Wahlin-Jacobsen, "Testosterone in women–the clinical significance," Lancet Diabetes Endocrinol, vol. 3, pp. 980–992, Dec. 2015.

[411] S. J. Fuller, R. S. Tan, and R. N. Martins, "Androgens in the etiology of alzheimer's disease in aging men and possible therapeutic interventions," J. Alzheimers. Dis., vol. 12, pp. 129–142, Sept. 2007.

[412] E. R. Rosario, L. Chang, E. H. Head, F. Z. Stanczyk, and C. J. Pike, "Brain levels of sex steroid hormones in men and women during normal aging and in alzheimer's disease," Neurobiol. Aging, vol. 32, pp. 604–613, Apr. 2011.

[413] D. K. Franco-Bocanegra, C. McAuley, J. A. R. Nicoll, and D. Boche, "Molecular mechanisms of microglial motility: Changes in ageing and alzheimer's disease," Cells, vol. 8, p. 639, June 2019.

[414] C. M. Karch, L. A. Ezerskiy, S. Bertelsen, Alzheimer's Disease Genetics Consortium (ADGC), and A. M. Goate, "Alzheimer's disease risk polymorphisms regulate gene expression in the ZCWPW1 and the CELF1 loci," PLoS One, vol. 11, p. e0148717, Feb. 2016.

[415] G. Novikova, M. Kapoor, J. Tcw, E. M. Abud, A. G. Efthymiou, S. X. Chen, H. Cheng, J. F. Fullard, J. Bendl, Y. Liu, P. Roussos, J. L. Björkegren, Y. Liu, W. W. Poon, K. Hao, E. Marcora, and A. M. Goate, "Integration of alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes," Nat. Commun., vol. 12, p. 1610, Mar. 2021.

[416] B. W. Henderson, E. G. Gentry, T. Rush, J. C. Troncoso, M. Thambisetty, T. J. Montine, and J. H. Herskowitz, "Rho-associated protein kinase 1 (ROCK1) is increased in alzheimer's disease and ROCK1 depletion reduces amyloid-β levels in brain," J. Neurochem., vol. 138, pp. 525–531, Aug. 2016.

[417] E. G. Gentry, B. W. Henderson, A. E. Arrant, M. Gearing, Y. Feng, N. C. Riddle, and J. H. Herskowitz, "Rho kinase inhibition as a therapeutic for progressive supranuclear palsy and corticobasal degeneration," J. Neurosci., vol. 36, pp. 1316–1323, Jan. 2016.

[418] M.-F. Guo, H.-Y. Zhang, Y.-H. Li, Q.-F. Gu, W.-Y. Wei, Y.-Y. Wang, X.-J. Zhang, X.-Q. Liu, L.-J. Song, Z. Chai, J.-Z. Yu, and C.-G. Ma, "Fasudil inhibits the activation of microglia and astrocytes of transgenic alzheimer's disease mice via the downregulation of TLR4/Myd88/NF-κB pathway," J. Neuroimmunol., vol. 346, p. 577284, June 2020.

[419] J. Chen, Z. Sun, M. Jin, Y. Tu, S. Wang, X. Yang, Q. Chen, X. Zhang, Y. Han, and R. Pi, "Inhibition of AGEs/RAGE/Rho/ROCK pathway suppresses non-specific neuroinflammation by regulating BV2 microglial M1/M2 polarization through the NF-κB pathway," J. Neuroimmunol., vol. 305, pp. 108–114, Apr. 2017.

[420] A. Kroiss, S. Vincent, M. Decaussin-Petrucci, E. Meugnier, J. Viallet, A. Ruffion, F. Chalmel, J. Samarut, and N. Allioli, "Androgen-regulated microRNA-135a decreases prostate cancer cell migration and invasion through downregulating ROCK1 and ROCK2," Oncogene, vol. 34, pp. 2846–2855, May 2015.

[421] A.-E. Roser, L. Tönges, and P. Lingor, "Modulation of microglial activity by Rho-Kinase (ROCK) inhibition as therapeutic strategy in parkinson's disease and amyotrophic lateral sclerosis," Front. Aging Neurosci., vol. 9, p. 94, Apr. 2017.

[422] A. M. Horstman, E. L. Dillon, R. J. Urban, and M. Sheffield-Moore, "The role of androgens and estrogens on healthy aging and longevity," J. Gerontol. A Biol. Sci. Med. Sci., vol. 67, pp. 1140–1152, Nov. 2012.

[423] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. del Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R. C. Lovering, B. Meldal, A. N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni, and H. Hermjakob, "The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases," Nucleic Acids Res., vol. 42, pp. D358–63, Jan. 2014.

[424] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub, "The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease," Science, vol. 313, pp. 1929–1935, Sept. 2006.

[425] A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W.-N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E. Root, B. Wong, and T. R. Golub, "A next generation connectivity map: L1000 platform and the first 1,000,000 profiles," Cell, vol. 171, pp. 1437–1452.e17, Nov. 2017.

[426] Y.-Y. Wang, H. Kang, T. Xu, L. Hao, Y. Bao, and P. Jia, "CeDR atlas: a knowledgebase of cellular drug response," Nucleic Acids Res., vol. 50, pp. D1164–D1171, Jan. 2022.

[427] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech: Theory Exp., vol. P10008, pp. 1–12, Mar. 2008.