Bridging between data and knowledge: New ways to better understand Alzheimer's Disease and Type 2 Diabetes comorbidity

Kumulative Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

REAGON KARKI

aus Kathmandu, Nepal

BONN, 2021

Angefertigt mit Genehmigung

der Mathematisch-Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. rer. nat. Martin Hofmann-Apitius

2. Gutachter: Prof. Dr. Diana Imhof

Tag der Promotions: 12.02.2021

Erscheinungsjahr: 2021

“Under everything just another human being”

– Eddie Vedder

Declaration

I herewith certify that this thesis is my own work, that I used only those sources and resources referred to in the thesis, and that I have identified citations as such.

_____

Reagon Karki

Acknowledgement

This thesis is an outcome of years of dedicated time, energy and effort I spent on doing science and research. I would like to thank everyone involved in this work and those who contributed.

My utmost gratitude to Prof. Dr. Martin Hofmann-Apitius for taking me under his wing and providing me the opportunity to contribute to the scientific community. I am thankful for his constant guidance, supervision and support throughout my PhD which has helped me overcome all of the difficult times along the way. Thank you very much.

I would like to thank my second supervisor, Prof. Dr. Diana Imhof, for her constructive comments and suggestions. I would also like to thank Prof. Dr. Thomas Schultz and Dr. Arijit Biswas for agreeing to be in the defense committee.

My special thanks to Dr. Alpha Tom Kodamullil for her support, guidance and valuable suggestions during my PhD. I learned a lot working with her and enjoyed my time in all of the projects we did together.

I am indebted to Alina Enns and Meike Knieps for assisting me with all the administrative works.

The working atmosphere at SCAI-BIO and bonding with my colleagues was full of positive vibes and energy. I am glad to be part of such a special group. Thank you all for that.

I am deeply grateful to Dr. Christine Robinson for helping me to improve the language and writing of my thesis.

I owe my deepest gratitude to my family and friends for their continuous encouragement and support all these years. Your love, care and blessings have helped me follow my dreams.

Lastly, I wish to thank my beloved deceased wife Dr. Shilva Kayastha (1987–2018) for being the light and mana of my life. Thank you for making me a betterman.

Abstract

The holistic understanding of biological phenomena enabled by systems biology forms the basis of current research. As opposed to the principle of classical biology where behavior of a biological system is explained by studying individual constituents, systems biology evaluates many constituents of a system simultaneously to explain how molecular processes influence higher level biological phenomena. The integration of knowledge acquired through published literature and data generated from direct experimentation is the most important feature of systems biology because it facilitates communication between these worlds and overcomes each other's weaknesses. Today, it has profound applications in a wide range of disciplines such as biomarker identification, drug discovery, network analyses and disease-mechanism identification.

In this thesis, using state-of-the-art frameworks and technologies of systems biology, we have performed a comparative analysis of disease-specific models to depict mechanism-centric comorbid association between Alzheimer's Disease (AD) and Type 2 Diabetes Mellitus (T2DM). We achieved this through two different methodologies where literature-based findings were validated with publicly available data and vice-versa. The findings from our first methodology illustrate cross-talk between several signaling pathways which eventually manifest characteristic features of AD and T2DM. Our findings provide a wider and global overview of previously suggested comorbidity between these diseases. Moreover, we have explored putative beneficial and harmful effects induced by Metformin, an FDA-approved T2DM drug which is considered a candidate repurposing drug for AD. With our second methodology, we have identified four pleiotropic genes to be involved in pathophysiological events of both AD and T2DM. Interestingly, these genes did not fall into the category of well-known genes of both diseases, suggesting a new mechanistic route to the comorbid association. In addition to the work that

explores AD-T2DM comorbidity, this thesis also focuses on a work that devises a new algorithm to enable quantification of disease mechanisms. The algorithm, named Candidate Mechanism Perturbation Algorithm (CMPA), was able to demonstrate that the intensity of impairment of causal mechanisms is different across spatial and temporal resolutions. Such an implementation opens up the possibility to generate a ranked and prioritized list of disease mechanisms.

Lastly, this thesis endorses understanding of disease mechanisms and network analysis by providing robust and reproducible workflows. The works presented here introduce methodologies to bring new insights about comorbid diseases and score disease-associated mechanisms.

Publications

- **Karki R**, Kodamullil AT, Hofmann-Apitius M. *Comorbidity Analysis between Alzheimer's Disease and Type 2 Diabetes Mellitus (T2DM) Based on Shared Pathways and the Role of T2DM Drugs.* Journal of Alzheimer's Disease, 60(2), 721-731. (DOI:10.3233/JAD-170440)

- **Karki R**, Kodamullil AT, Hoyt CT, Hofmann-Apitius M. *Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation amplitude (CMPA) algorithm.* BMC bioinformatics, 20(1), 1-8. (DOI: 10.1186/s12859-019-3101-1)

- **Karki R**, Madan S, Gadiya Y, Domingo-Fernández D, Kodamullil AT, Hofmann-Apitius M. *Data driven approaches in understanding comorbidity between Type 2 Diabetes Mellitus (T2DM) and Alzheimer's Disease (AD).* Journal of Alzheimer's Disease, (Preprint), 1-9. (DOI: 10.3233/JAD-200752)

- Fluck J, Madan S, Ansari S, Kodamullil AT, **Karki R**, Rastegar-Mojarad M, Catlett NL, Hayes W, Szostak J, Hoeng J, Peitsch M. *Training and evaluation corpora for the extraction of causal relationships encoded in biological expression language (BEL)* Database Journal, 2016. (DOI: 10.1093/database/baw113)

- Kodamullil AT, Iyappan A, **Karki R**, Madan S, Younesi E, Hofmann-Apitius M. *Of Mice and Men: Comparative Analysis of Neuro-Inflammatory Mechanisms in Human and Mouse Using Cause-and-Effect Models.* Journal of Alzheimer's Disease, 59(3), 1045-1055. (DOI: 10.3233/JAD-170255)

- Emon MA, Kodamullil AT, **Karki R**, Younesi E, Hofmann-Apitius M. *Using Drugs as Molecular Probes: A Computational Chemical Biology Approach in Neurodegenerative Diseases*. Journal of Alzheimer's Disease, 56(2), 677-686. (DOI: 10.3233/JAD-160222)

- Domingo-Fernández D, Kodamullil AT, Iyappan A, Naz M, Emon MA, Raschka T, **Karki R**, Springstubbe S, Ebeling C, Hofmann-Apitius M. *Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): a web server for mechanism enrichment* Bioinformatics, 33(22), 3679-3681. (DOI: 10.1093/bioinformatics/btx399)

- de Jong J, Emon MA, Wu P, **Karki R**, Ahmad A, Froehlich H. *Deep learning for clustering of multivariate clinical patient trajectories with missing values*. GigaScience, 8(11), giz134. (DOI: 10.1093/gigascience/giz134)

- Sahay A, Sood M, **Karki R**, Emon MA, Vrooman H, Hofmann-Apitius M, Froehlich H. *Realistic Simulation of Virtual Multi-Scale, Multi-Modal Patient Trajectories using Bayesian Networks and Sparse Autoencoders*. Scientific reports, 10(1), 1-14. (DOI: 10.1038/s41598-020-67398-4)

- Domingo-Fernández D, Baksi S, **Karki R**, et al. *COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology*. BIOINFORMATICS, accessible online, 25 September 2020. (DOI: 10.1093/bioinformatics/btaa834)

Contents

List of Abbreviations

AD	Alzheimer’s Disease
PD	Parkinson’s Disease
NDD	Neurodegenerative Disease
T2DM	Type 2 Diabetes Mellitus
T3DM	Type 3 Diabetes Mellitus
NFT	Neurofibrillary Tangle
RCR	Reverse Causal Reasoning
BN	Bayesian Network
NPA	Network Perturbation Amplitude
CMPA	Candidate Mechanism Perturbation Amplitude
GWAS	Genome Wide Association Studies
BEL	Biological Expression Language
LD	Linkage Disequilibrium
SNP	Single Nucleotide Polymorphism

Chapter 1

Introduction

1. Revolutionizing translational research with integrative approaches

Classical biology, also known as reductionist biology, was the basis for all biological research for over two centuries, and is driven by single-molecule experiments and small-scale modeling. It has contributed significantly to our understanding of biological systems, via examination of properties of individual components of those systems. In fact, the identification of many molecules, genes and biological processes is attributable to years of research that followed the principles of classical biology [1]. Despite its many contributions, classical biology has not been able to adequately describe complex diseases. Classical approaches use observed disease phenotype or expert opinions to suggest a hypothesis. The results of this process, however, are often misguided, because the hypotheses may reflect unrecognized bias or incomplete insights about the relevant disease. Moreover, a number of limitations of classical biology have hindered swift progress in the field of molecular biology. First, classical approaches concentrate on understanding physiochemical and biological properties of selective molecules. Second, the process is not time-efficient and has many assumptions and constraints. Third, it lacks a philosophical framework to investigate systems as a whole, that is, the interaction and interplay of components that affect behavior of the system. The repeated failure of clinical and drug trials, including FDA-approved drugs to treat Alzheimer's Disease (AD), provides a costly reminder of the limitations of classical biology [2]. The case of AD is a particularly stark demonstration of the need for non-classical approaches. Over 100 years of research and billions of dollars of investment have made seemingly little change to its status as "incurable". The challenge of decoding the true etiology of AD stems from the fact that it is a multi-factorial and complex

neurological disorder. Progress in understanding such disorders requires a holistic understanding and identification of concurrent cross-talk of its pathological, demographical, epidemiological and molecular aspects. Such a perspective is potentially offered by systems biology approaches, which bridges the gap between multi-scale biomedical data and knowledge ranging from genotype to phenotype in diseases.

The concept of systems biology was first described in the early 1920s by Ludwig von Bertalanffy [3] and in the late 1960s by Mijajlo Mesarovic [4]. However, the real implementation of systems biology approaches only became possible with technological advancements and developments of the 1990s, and such approaches have since become standard**.** And within a short span of time, systems biology has revolutionized molecular biology and all of its subfields. The integration of data-driven and knowledge-driven approaches has been the essence of systems biology. These approaches complement each other, and so, their integration facilitates more powerful scientific findings and interpretations. While data-driven approaches are capable of generating bias-free hypotheses, knowledge-driven approaches provide explanations of molecular interactions and downstream molecular events, which data-driven methods cannot. Since the early 1990s, especially after the commencement of the Human Genome Project (HGP), there has been a dramatic increase in collection of high-throughput omics data including metabolomics, genomics, transcriptomics and proteomics [5]. Today, advanced technologies and computer systems facilitate the acquisition of heterogeneous, multi-scale and multi-modal data, often referred to as "big data". Concurrently, numerous methodologies and algorithms have been developed for the analysis and interpretation of this data. Meanwhile, knowledge-driven approaches have generated insight about biological entities and their functions. This knowledge is published as literature, and has kept pace with data. However, the process of publishing scientific literature is a slow one, and this knowledge is typically an outcome of conventional practices of hypothesis generation. This

is vulnerable to the same risk of bias as classical biology. The individual limitations of data- and knowledge-driven approaches motivate the need for something that improves them. Interestingly, these approaches are complementary, in the sense that each seems to overcome a weakness of the other. For instance, data-driven approaches can detect a signal, but not generate insight about its functional context. That insight can, however, be provided by the prior knowledge offered by knowledge-driven approaches. In this context, a study by Khanna et al. (2018) has generated Bayesian Networks (BNs) to uncover dependencies across multi-scale patient level data where patients transition from normal state to the state of mild cognitive impairment [6]. The authors have used knowledge assembly to reconstruct potential biological mechanisms embedding dependent features that influence this transition. Likewise, a finding from knowledge-driven approaches requires validation with data. This has been demonstrated in a study that performs a comparative analysis of neuro-inflammatory mechanisms in human and mouse. The concordance of gene expression patterns between the two species is validated using independent gene expression datasets [7].

This potential to enable interoperability and communication between the worlds of data and knowledge marks one of the points of true potential of systems biology. Its holistic approach to understanding complex biological systems, modeling metabolic networks, cell signaling networks and disease models, is why it forms the foundation of most current research. Moreover, systems biology has found application in a wide range of sub-disciplines, such as chemical synthesis [8], biomarker identification [9], drug discovery [10], personalized and precision medicine [11], immunology and vaccination [12][13], proteome profiling [14][15], *in silico* simulations with cell-free-systems (CFS) [16], network-based analyses [17][18] and disease-mechanism identification [19]. Lately, there are even advanced methodologies that have influenced paradigm shift in systems biology as progresses have been made beyond data-knowledge integration and

interoperability. A hybrid data and knowledge driven framework, namely CLEP, has enabled generation of patients which can be further used for clustering and classification [20]. The framework uses several machine learning algorithms that allow accurate classification between cognitively impaired patients and healthy participants.

In this thesis, we present novel systems biology approaches in the domains of network-based analysis and disease-mechanism identification. In particular, by creating and comparing disease specific biological networks, works that identify shared molecular mechanisms of co-occurring diseases are demonstrated here. This sort of approach provides mechanistic insights about co-occurring diseases, which are not possible through classical assessment techniques, as they examine diseases from clinical findings and observations. Moreover, although several mechanisms are suggested to be involved in underlying disease etiology, it is difficult to identify the most crucial and important mechanisms. A strategy to score disease mechanisms will resolve this issue. A work describing a new algorithm that assesses intensity of impaired disease mechanisms is presented here. In the remainder of this introduction, we will provide a summary of the state-of-the-art technology and research in these domains.

2. Comorbidity

The term comorbidity refers to the simultaneous occurrence of two or more medical conditions in an individual. It was coined by A.R Feinstein in 1970 and since then it has been considered as a separate domain of scientific research in many branches of medicine [21]. Based on the association between co-occurring conditions, comorbidity is considered to be either non-dependent or dependent. Non-dependent comorbidity describes the co-occurrence of conditions due to random probability, suggesting no association between the diseases. Dependent comorbidity, on the other hand, suggests that the conditions are related in some way or that one condition is a risk factor for another condition. As opposed to non-dependent comorbidity, conditions exhibiting dependent comorbidity are assumed to share underlying disease pathways and mechanisms.

2.1. Prevalence and causes of comorbidity

The prevalence of comorbidities varies among patient subpopulations and different families of diseases. Irrespective of its type, comorbidity affects mortality, functional status and quality of life of patients. Moreover, it influences their prognosis, medication, treatment strategy and overall handling [22]. A population-based study of 8000 patients by Roca et al. (2009) showed that more than 30% of the patients were diagnosed with coexisting psychiatric disorders [23]. As shown in Figure 1, 30.1% of the patients were diagnosed with concurrent mood disorders and somatoform disorders. Similarly, 30.6% of patients were diagnosed with mood disorders and anxiety disorders. Moreover, a total of 11.5% of the patients were diagnosed with all the 3 conditions.

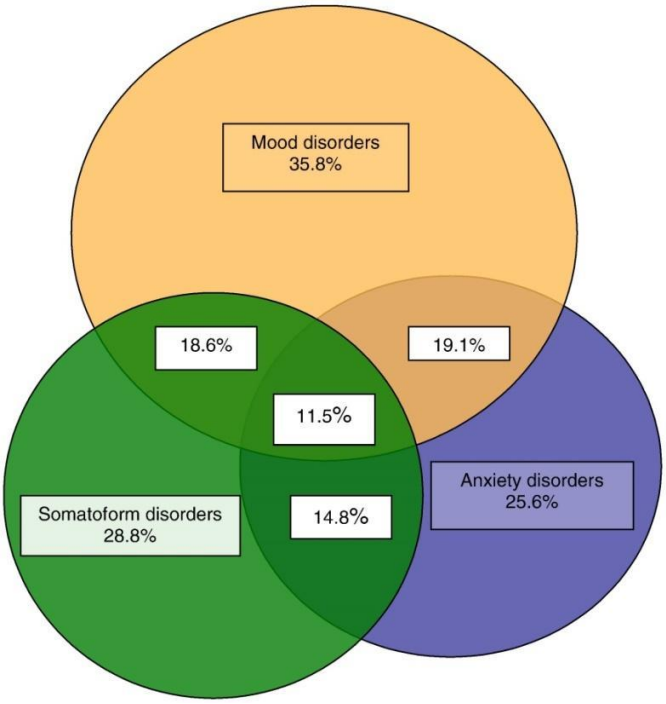

Figure 1: A Venn diagram depicting proportions of prevalence of comorbid conditions in a Spanish population diagnosed with mood disorders, somatoform disorders and anxiety disorders. (Source: Roca et al. (2009) [23])

Likewise, the Alzheimer's Association (www.alz.org) (Figure 2) has demonstrated that the number of dementia patients with additional co-existing medical conditions is more than the number of dementia patients with no additional medical condition. The comorbid conditions were congestive heart failure, chronic obstructive pulmonary disease, chronic kidney disease, coronary artery disease, stroke, diabetes and cancer, and were diagnosed in patients aged 65 years and older. The costs of medication and health care service of dementia patients with comorbid conditions was reported to be higher than costs for dementia patients with no comorbid condition.

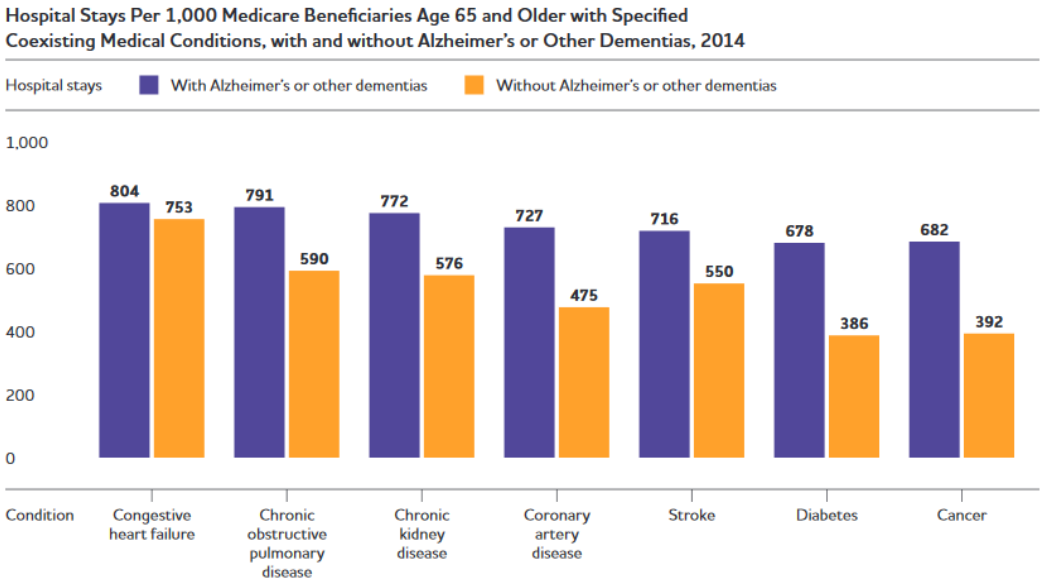

Figure 2: Bar plots comparing occurrence of comorbid conditions per 1000 US patients of age 65 or older diagnosed with and without AD. The purple bars represent patients with AD and yellow bars represent patients without AD. This comparison indicates that prevalence of comorbid conditions in AD patients is higher than in patients without AD. (Source: www.alz.org)

In another cohort dataset of patients aged 65 or older (i.e., The Health Improvement Network (THIN)), 77% of patients with dementia were diagnosed with one or more comorbid conditions (Figure 3). Additionally, the number of patients with dementia and co-existing comorbid condition(s) was found to be greater than the number of patients with the same comorbid condition(s), but without dementia. The most prevalent comorbid conditions were hypertension and diabetes, both of which are considered to be a risk factor for dementia.

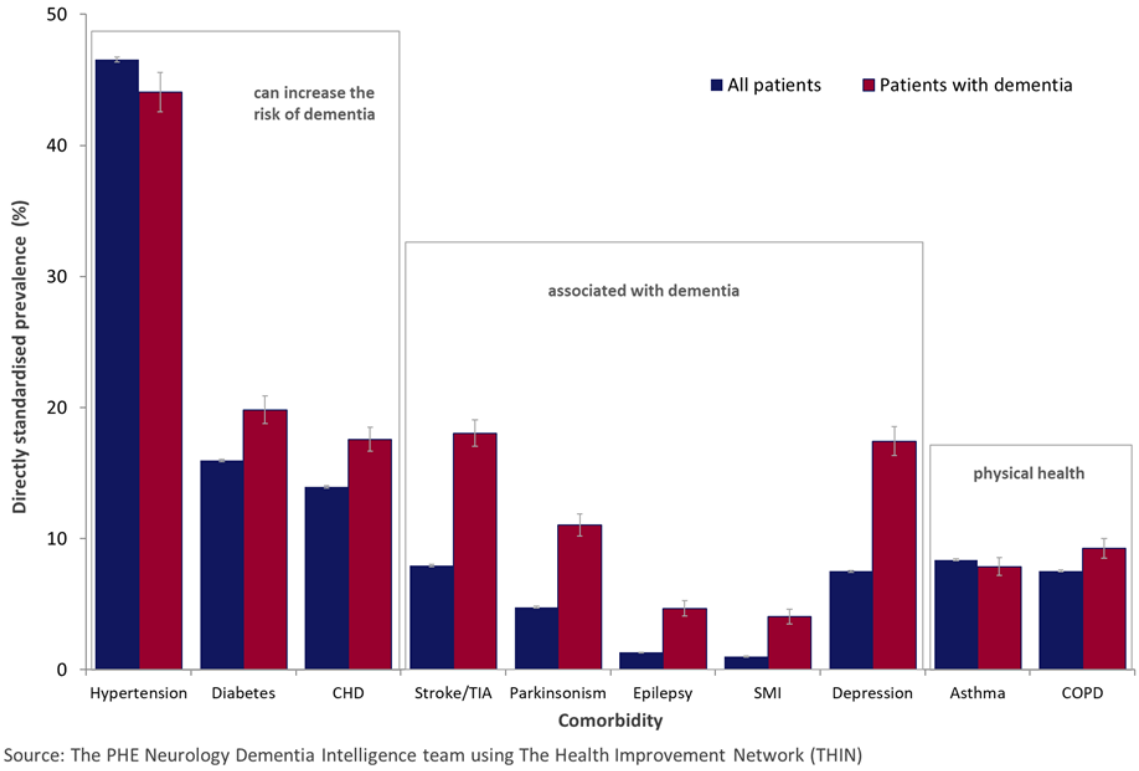

Figure 3: Bar plots comparing occurrence of comorbid conditions in UK patients of age 65 or older diagnosed with AD against all patients diagnosed with different diseases. The blue plots represent patients with AD and red plots represent entire group of patients with different diseases. This comparison indicates that prevalence of comorbid conditions in AD patients is higher when compared to all the patients with various diseases. (Source: www.gov.uk)

While a full understanding of the causes of comorbidity is still lacking, a number of heterogeneous factors are associated with comorbidity, including demographics, lifestyle, social and physical environment and health care. In recent years, there has been growing interest in the scientific community as comorbidity (especially dependent type) is speculated to exist at multi-scale biology of organs, pathways, cellular processes and genetics. Sánchez-Valle et al. (2020) report on the findings that comorbid conditions are influenced by the physical closeness of organs. In other words, a patient with a disease in the digestive system is more likely to develop another disease associated with the digestive system. Through a Disease Molecular Similarity Network (DMSN), they identified higher overlap of similarity interactions between diseases of

the digestive system, diseases of skin and cancer [24]. A study by Ko et al. (2016) created molecular interaction networks by retrieving disease-associated genes and underlying molecular pathways from available databases to identify comorbid diseases. The study speculates comorbidity between Diabetes Mellitus and Ankylosing spondylitis and other inflammatory spondylopathies on the basis of identified common mechanisms such as interleukin-10 receptor binding, regulation of immune response and response to insulin [25]. The high prevalence of this comorbidity has been reported in a study based on an Asian population [26]. Similarly, by identifying a shared mechanism, Hoyt et al. (2018) have attempted to explain previously suggested comorbidity between AD and Epilepsy by depicting the putative role of GABA receptors and consequent GABAergic pathway as the underlying comorbid etiology [27]. In recent years, besides identification of shared genes, biological processes and pathways, the influence of genetics has also been assessed in deducing comorbid inferences. Such studies have undertaken efforts to identify pleiotropic genes and even their variants associated with downstream protein modifications, function and consequent phenotypes that are shared between comorbid conditions. In this regard, Tomblin et al. (2012) have suggested involvement of four pleiotropic genes (i.e., KIA0319, DRD4, DAT1 and BDNF) in comorbidity between Communication Disorder (CD) and Attention Deficit Hyperactivity Disorder (ADHD). While procedural learning and declarative learning are cognitive endophenotypes of CD, and executive function and procedural learning are of ADHD, the authors rationalize comorbidity between these by illustrating deficits in all of these endophenotypes caused by the aforementioned pleiotropic genes [28]. Another study by Naz et al. (2017) performed functional assessment of pleiotropic variants to explain the possible stress-induced comorbidity between AD and Parkinson's Disease (PD). The study identified the rs1800547 variant in the MAPT gene of AD and PD patients as

contributing to the characteristic features of those diseases. In addition, the authors also report that the rs393152 variant in CRHR1 adds to the underlying comorbid etiology [29].

2.2. Overview of Alzheimer's Disease

AD, the most common form of dementia, is a progressive neurodegenerative disease which affects the aging population. The disease is characterized by loss of structure and function of neurons in the brain, resulting in loss of memory, complications in executive functions, inability to communicate properly and perform regular day to day activities. The accumulation of amyloid beta peptides and neurofibrillary tangles (NFTs) resulting from abnormally processed APP and misfolded MAPT, respectively, are the most remarkable clinical hallmarks of AD [30][31]. The hypotheses explaining the exhibition of these hallmarks are widely known as amyloid hypothesis and tau hypothesis. In addition to theses, the etiology of AD is associated to other speculations, such as decreased acetylcholine synthesis in signal transduction (i.e., cholinergic hypothesis) [32], improper functioning of the blood-brain barrier (i.e., neurovascular hypothesis) [33], increased inflammatory markers (i.e., inflammation hypothesis) [34], reduced glucose metabolism (i.e., glucose hypothesis) [35], and perturbed cellular bio-metals homeostasis [36]. AD is presumed to be a genetic disease as mutations in the genes APP, PSEN1, APOE and TREM2 are known to increase the risk of developing AD at least threefold [37][38]. Even more, the carriers of the APOE ε4 isoform are considered to be the most vulnerable risk group as studies have shown this to increase the risk of AD by fifteen times [39]. The other reported risk factors in AD are smoking, lifestyle, diet, nutrition and environmental factors [40][41][42]. The inefficacy of available drugs targeting these hypotheses suggests that AD is a complex neurological disorder and demands more research to identify potential new hypotheses and drug targets.

2.3. Overview of Type 2 Diabetes Mellitus

T2DM is a metabolic disorder characterized by inability of cells to effectively utilize insulin required for conversion of glucose to glycogen (glycogen synthesis) which thereby increases blood glucose levels (hyperglycemia). The prevalence of T2DM is mostly found in the middle-aged and older population, with about 400 million people affected worldwide [43]. Over the years, the understanding of T2DM has changed from being a simple disorder of the pancreas to a complex multi-system disorder, with organs such as liver, muscles, kidney, gut and brain known to be affected by the disease [44]. The etiology of T2DM is widely attributed to pre-existing obesity, lifestyle, medical conditions and genetics. Population based studies have shown that 30-100% of individuals with obesity developed T2DM [45]. Moreover, individuals with a lifestyle of smoking, limited physical activity and consumption of a high sugar diet are more likely to get the disease [46][47][48]. Prolonged use of medications such as glucocorticoids, thiazides, antipsychotics and statins are also reported to increase the risk of T2DM in several independent studies [49][50]. While a few studies investigate a dysfunctional monogenic (i.e., single gene) form of T2DM [51], a number of studies suggest that the disease is polygenic in nature. Genome Wide Association Studies (GWAS) have identified 86 single nucleotide polymorphisms (SNPs) in about 35 genes that contribute to the pathophysiology of T2DM [52][53]. Among these, TCF7L2 is the most prominent gene, because it regulates production of glucagon (GCG), a gene which enhances insulin secretion. In another GWAS, SNPs were identified in TBC1D4, a gene known to enhance insulin-mediated uptake of glucose [54]. The overall population had a tenfold increase in risk of developing T2DM, compared to those without this mutation. As compared to AD, the etiology of T2DM is well-understood, reflected in the existence of several effective therapies to treat the disease.

2.4. The comorbid story of AD and T2DM

The detection of insoluble IAPP, which resembles the toxic amyloid beta peptides of AD, in the pancreas of diabetic patients in the late 1980s was one of the first hints that suggested a pathological association between AD and T2DM [55]. Since then it has been the focus of massive interest in the scientific community to unravel existing links between the neurological disorder AD and the metabolic disorder T2DM. Associations between AD and T2DM have now been established from different perspectives, such as epidemiology, neuroimaging, shared genetics and shared pathophysiology. Several different types of potential associations between AD and T2DM have been proposed, with each of which has its own implications for the understanding of the diseases. The possible associations include: a generic but close relationship, a co-dependent association, that T2DM is a risk factor for AD and most strongly that AD is, in fact, a Type 3 Diabetes Mellitus (T3DM) [56]. Irrespective of the nature of the association, the findings make it clear that the co-occurrence of these diseases is more than just by chance, and they are, in fact, comorbid in nature.

The Rotterdam Study of the 1990s was one of the first epidemiological studies investigating the impact of T2DM as a risk factor for AD. It paved the way for many future cohort-based studies with the same goal [57]. The study identified T2DM as the strongest risk factor for AD. The prevalence of AD in T2DM patients was found to be twice as high as in the control patients. In a study by Barbagallo and Dominguez, T2DM patients also had a two-fold risk of developing AD compared to healthy controls [58]. Similarly, longitudinal population-based studies have also revealed that T2DM patients are at significantly higher risk of developing AD compared to individuals without T2DM [59][60]. Likewise, several studies have reported that the APOE ε4 isoform in T2DM patients increases the risk of AD, compared to non-carriers of APOE ε4

isoform [61][62][63]. A study by Li et al. (2015) summarizes seventeen epidemiological studies that suggest a comorbid association between AD and T2DM [64].

The finding that T2DM patients are confronted with cognitive decline as a result of reduced hippocampal and whole brain volume implies that it is more than just a metabolic disorder [65]. These observations resemble the typical pattern of brain atrophy and consequent impairment of cognition evident in AD. In this context, several other neuroimaging studies have reported cognitive impairment and its correlation with reduced volumes of gray matter, white matter and hippocampus in T2DM patients [66][67]. While epidemiological and neuroimaging studies are based on observations and can elucidate on correlations or dependencies between considered features of the studies, they are still incapable of giving insights about pathophysiological events underlying comorbid conditions. Therefore, it is of utmost importance to understand comorbidity at the genetic and molecular levels. This will identify and provide understanding of genetic variants, dysfunctional genes, impaired biological processes and pathways involved in comorbid conditions. The molecular processes linking AD and T2DM as comorbid conditions are explained as follows:

2.4.1. Insulin signaling

The prevalence of impaired insulin signaling, a characteristic feature of T2DM, in the brains of AD patients is considered to be strong evidence to why AD and T2DM are comorbid. The normal regulation of insulin signaling balances glucose metabolism, as insulin increases uptake of glucose in fat and muscle cells. However, these cells are incapable of utilizing available insulin in T2DM, a condition referred to as insulin resistance. As a consequence, patients experience hyperglycemia, abnormally high blood glucose levels. The impairment of insulin signaling which leads to insulin resistance and hyperglycemia in T2DM is known to be caused by interference by pro-inflammatory cytokines (TNF and interleukins) and abnormal phosphorylation of insulin

growth factor receptors. In AD, on the other hand, insulin resistance is directly attributed to accumulation of amyloid beta peptides and hyper-phosphorylated MAPT yielding NFTs, both of which are clinical hallmarks of AD (Figure 4) [68].

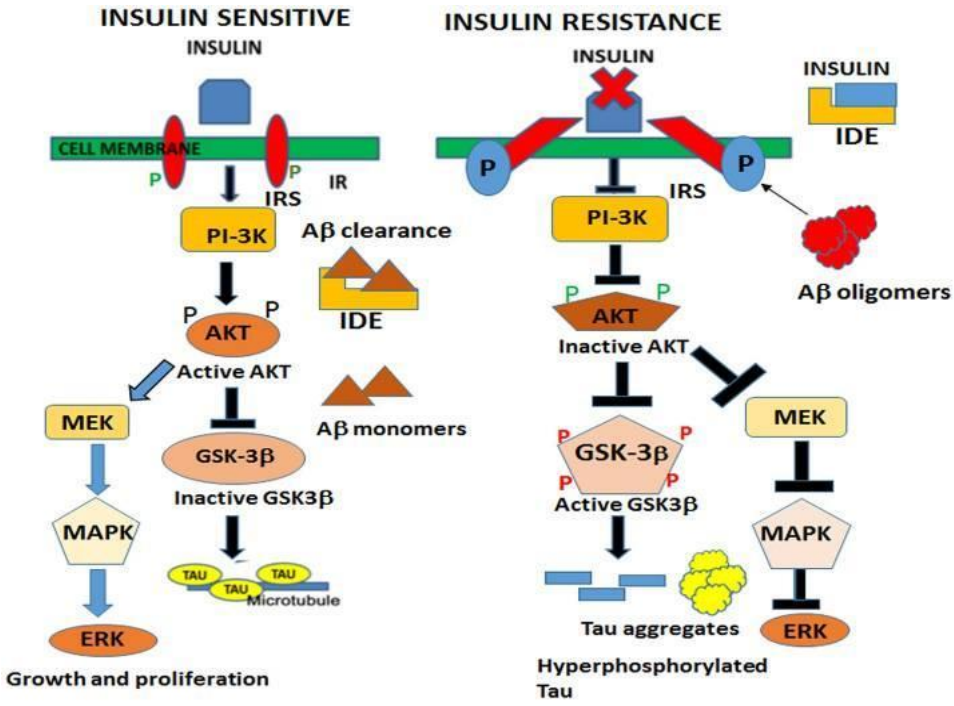

Figure 4: Molecular interactions involved in insulin signaling. The interactions on the left represent downstream effects of a normally regulated insulin signaling (i.e., insulin sensitive). This leads to 1) cell growth and proliferation via activation of ERK and 2) inactivation of GSK3B which prevents phosphorylation of MAPT. The interactions on the right represent consequences of impaired insulin signaling (i.e., insulin resistance). This impairment, caused by amyloid beta peptides, inhibits binding of insulin with its receptors, thereby hindering downstream inactivation of GSK3B. This causes hyperphosphorylation of MAPT, which eventually leads to formation of neurofibrillary tangles (NFTs). Moreover, cell growth and proliferation is inhibited as a result of inhibited ERK. (Source: Chatterjee et al. (2018) [68])

In particular, insulin signaling involves GSK3B, an enzyme whose kinase activity is reckoned to be the most deleterious in AD because it actively induces hyper-phosphorylation of MAPT. Moreover, several studies have identified defects in insulin signaling as a cause of accumulation of amyloid beta peptides. This is explained by observations that NEP and IDE are reduced in

impaired insulin signaling, both of which are amyloid beta degrading enzymes [69]. Interestingly, amyloid beta peptides are known to interfere with insulin signaling by abnormally phosphorylating IRS1 through the JNK pathway [70]. This suggests that there is a feed-forward loop where amyloid beta peptides impair insulin signaling, which in turn results in reduced clearance of amyloid beta peptides. A review by Chatterjje et al. (2018) provide a critical assessment of how T2DM-related pathological traits such as hyperinsulinemia, insulin resistance and hyperglycemia influence the presentation of AD-related pathological traits such as deposition of amyloid beta-peptides and NFTs [68]. Besides these, the role of insulin in improving memory and cognition has been identified through animal and human-based studies [71]. For this reason, intranasal insulin has been recommended as a potential therapy for prevention of AD [72].

2.4.2. Inflammation

Inflammation has been proposed as a possible mechanistic link between AD and T2DM, based on the occurrence of peripheral and neuronal inflammation in T2DM and AD respectively (Figure 5) [73]. The review hypothesizes that peripheral inflammation induced T2DM, identified by increased levels of inflammation mediators in the vascular system, increases amyloid beta peptides in the brain through stimulation of advanced glycation end products (AGEs). The amyloid beta peptides then further exacerbated T2DM severity and induced synaptic dysfunction in AD. On the other hand, neuronal inflammation, identified by elevated levels of inflammatory mediators in the brain, directly induced synaptic dysfunction.

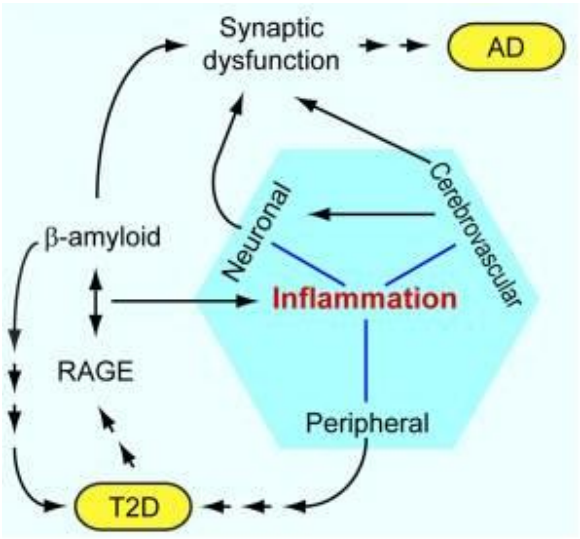

Figure 5: A chain of interactions depicting the role of peripheral inflammation in T2DM which eventually contributes to synaptic dysfunction in AD. Additionally, the prevalence of neuronal and cerebrovascular inflammation can directly influence synaptic dysfunction in AD. (Source: Han et al. (2010) [73])

The previous section about insulin signaling already gives a hint about involvement of inflammatory genes in T2DM. In this regard, animal models mimicking T2DM have shown increased levels of pro-inflammatory cytokines including TNF, IL1B, IL2 and IL6 [74]. Likewise, elevated levels of IL6 and CRP have been reported in human-based studies [75][76]. This increase in pro-inflammatory markers is known to be mediated by hyperglycemia induced reactive oxygen species [77]. Additionally, hyperglycemia-induced NFKB is reported to trigger inflammation and apoptosis in adipose tissue [78]. The inflammation that occurs in T2DM promotes upregulation of kinases such as IKBKB and MAPK8, both of which impair insulin signaling by abnormally phosphorylating IRS1 at Serine 307 [79]. By contrast, in AD, increased levels of inflammatory markers is considered to be an early sign of AD pathogenesis. In this regard, the genes A2M, CLU and CD74 were identified by independent studies [80][81]. Interestingly, unutilized insulin in cells (i.e., hyperinsulinemia), which is usually associated with

T2DM, is reported to increase pro-inflammatory cytokines such as IL1B, IL6 and TNF in AD [82]. As a consequence, insulin resistance in the brain takes place as the cytokines phosphorylate and activate IRS1. Moreover, these cytokines have also been identified to co-localize with amyloid beta peptides in AD, triggering tissue damage and neuronal death [83][84]. In fact, inflammation is speculated to be a downstream effect in the amyloid hypothesis, as amyloid beta peptides activate microglia which provoke a pro-inflammatory cascade. This results in production of cytokines, chemokines and reactive oxygen species, leading to loss of neuronal cells [85][86]. Likewise, inflammation has also been associated with the tau hypothesis in AD, from the findings that IL1B induces MAPT phosphorylation via MAPK pathway [87]. These lines of evidence clearly imply occurrence of inflammation in both T2DM and AD. In addition, it can be inferred that inflammation is the link connecting pathophysiological events of T2DM and AD.

2.4.3. Synaptic dysfunction

The loss of synapses and neuronal cell death in the cortex and hippocampus is considered to be the basis of cognitive decline and memory loss in AD [88][89]. Several studies have held amyloid beta peptides responsible for this synaptic dysfunction and cell death [90][91]. The accumulated amyloid beta peptides reduce the activity of AMPA and NMDA receptors, thereby impairing glutamatergic transmission in signal transduction [92]. Likewise, evidence of localization of phosphorylated MAPT in synapses of the brains of AD patients has been reported which disrupt the normal interaction between MAPT and FYN required for regulating NMDA receptor activity [93]. Moreover, animal studies have shown that abnormal phosphorylation of MAPT correlates with decreased levels of synaptic markers such as DLG4, SYN1, GRIN1 and GRIA1 implying synaptic dysfunction [94][95]. An independent study by Tackenberg et al. (2009) suggests that synaptic dysfunction is mediated by amyloid beta peptides in the early stage of AD whereas cell death is induced by abnormal MAPT phosphorylation in late stage AD [96].

On the other hand, patients with T2DM are also known to experience synaptic dysfunction. This may be because insulin signaling plays a critical role in synapse formation and neuronal plasticity [97][98][99]. The impairment in synaptic transmission has been depicted through a diabetic animal-based study where altered AMPA and NMDA receptor activities were observed [100]. In another such study, insulin prevented amyloid beta peptide-induced synaptotoxicity and long term potentiation (LTP) impairment [101][102]. Furthermore, GABA and glutamate, both of which are involved in synaptic transmission, were found to be reduced in hyperglycemic environments [103]. In contrast to this finding, transgenic mice in a hypoglycemic environment suffered from impaired memory, decreased LTP and cell death. However, this is in concordance with the notion that T2DM patients are vulnerable to cognitive decline as a consequence of hypoglycemia induced by prescribed drugs [104]. Taken together these findings indicate that normal regulation of insulin signaling is crucial for maintaining a healthy cognitive function.

2.4.4. Autophagy

The clearance of misfolded protein aggregates is regulated by a process called autophagy. It triggers systematic degradation and recycling of cellular components which, when accumulated, induce oxidative stress-induced cell death [105]. In the context of AD, insufficient regulation of autophagy is accounted for accumulation of deleterious amyloid beta peptides and NFTs. In T2DM, this inhibition of autophagy occurs through insulin resistance-impaired MTOR signaling [106][107]. A study by Nixon et al. (2007) has identified formation of immature autophagic vacuoles in AD brain biopsies of neocortical regions, suggesting impaired autophagy in AD [108]. This loss in autophagic activity is supported by a study which identified reduced activity of ATG7 and ATG5, both of which are crucial for regulation of autophagy [103]. Likewise, reduced autophagy and increased aggregation of NFTs was observed in a ATG7 knockout mice model, suggesting a crucial role of ATG7 in clearance of amyloid beta peptides and NFTs [109].

Similarly, ineffective clearance of amyloid beta peptides in AD is attributed to reduced levels of BECN1, which results in insufficient autophagy [110]. By contrast, in T2DM, insulin resistance results in oxidative stress, which further damages intracellular organelles such as endoplasmic reticulum and mitochondria. Since these organelles are involved in the activation of autophagy, degradation and clearance of misfolded proteins around them is affected [111]. Additionally, mouse-based studies have shown impairment in glucose intolerance and reduction in insulin secretion in environments of suppressed autophagy [112][113]. A rat model mimicking AD and T2DM has demonstrated accumulation of polyubiquitinated MAPT in neurons resulting from decreased SQSTM1, a protein responsible for degradation of ubiquitinated proteins via autophagy. In another similar study, reduced levels of autophagic markers such as ATG7 and MAP1LC3B were observed in the cerebral cortex and hippocampus of mice. As a consequence, defects in protein clearance and accumulation of autophagosomes were evident [114]. Taken together, impaired autophagy seems to be persistent in both AD and T2DM, and hence can be considered as a shared molecular process in the comorbid link between the diseases.

2.4.5. Lipid homeostasis

APOE is a protein primarily synthesized in the liver and brain, and plays a crucial role in transporting cholesterol and regulating lipid homeostasis [115][116][117]. Of all known risk factors, the APOE e4 isoform is considered to be the most significant risk factor for early and late onset of AD, as it results in a several fold increase in the risk of developing AD [118][119]. It is also known to increase accumulation of amyloid beta peptides and impair pathways responsible for their clearance [120]. In this regard, an independent study has postulated that lower binding affinity of the APOE e4 isoform with amyloid beta peptides could be the reason that clearance of amyloid beta peptides is perturbed in AD [121]. Moreover, the isoform carriers are vulnerable to oxidative stress and cholinergic dysfunction [122]. Interestingly, the presence of the APOE e4

isoform is found to be negatively correlated with IDE, an enzyme capable of degrading amyloid beta peptides [123]. In T2DM, on the other hand, hyperglycemia is known to cause dyslipidemia, as characterized by low levels of high-density lipid (HDL) cholesterol and high levels of triglycerides [124]. The prevalence of dyslipidemia in T2DM is also reported in several other studies [125][126]. In T2DM patients with AD and the APOE e4 isoform, reduced HDL cholesterol level and increased total cholesterol and triglyceride levels were observed compared to other groups, suggesting that the APOE e4 isoform was responsible for elevated blood lipid levels [127]. In another study, T2DM patients with the APOE e4 isoform were found to exhibit severe AD pathology (i.e., accumulation of neurite plaques and NFTs) when compared to T2DM patients without the e4 isoform [128]. To conclude, both AD and T2DM manifest impaired lipid homeostasis, and it can thus be considered yet another strong biological link between the diseases.

3. Algorithms in network-based analyses

Biological networks in general can be understood as a collection of inter-connected biological entities and can refer to various kinds of networks, such as protein-protein interaction (PPI) networks, metabolic networks, genetic interaction networks, cell signaling networks, gene co-expression networks and BNs. Some of these networks such as PPI and metabolic networks are outcomes of systematic mining of published literature. Interactions from such networks are further organized and re-arranged to demonstrate regulation of pathways, which eventually form the basis of canonical pathway databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome [129][130]. The issues of lack of disease-specific knowledge and limited coverage of depicted knowledge in these databases have been addressed with disease-specific databases and knowledge assemblies such as AlzPathway [131], Parkinson Disease Map [132] and NeuroMMsig [133]. The other types of networks are outcomes of analysis of omics data or

disease-associated clinical and demographic features. While gene co-expression networks are created with co-expressed genes which are inter-connected and clustered together, BNs are composed of interconnected nodes representing conditional dependencies between the measured features. Variational Autoencoder Modular Bayesian Network (VAMBN) is a good example of application of BN networks where the network model precisely reflects expected causal relationships in the multi-scale patient level data [134]. Such relationships between the data variables are abstract representations of correlated and dependent variables. They bear the potential to be rationalized with mechanistic biological interactions allowing cross-talk between the data and knowledge world of diseases [6]. In this thesis, we refer to networks as literature-derived products, prior biological knowledge or knowledge assemblies.

The interpretation of high-throughput omics data is aided by prior biological knowledge, as it provides inference about the activity of relevant biological entities, functional roles and downstream pathways. Gene set enrichment analysis (GSEA) is a classical approach where differentially expressed genes are mapped to signaling pathways regulated by predefined sets of genes. This overlap of differentially expressed genes with predefined sets of genes is assessed statistically by calculated confidence scores. Over the years, a number of algorithms have been devised to improve the understanding of omics data. One such algorithm is Pathifier [135] which takes into account expression profiles of tumor samples to score and identify dysregulated pathways from KEGG [129], BioCarta [136] and Pathway Interaction Database (PID) [137]. The algorithm, capable of transforming gene-level information into pathway-level information, identified CXCR3 mediated signaling and oxidative phosphorylation to be associated with survival of patients. However, recent studies have exposed some limitations of canonical pathways by identifying discrepancies and inconsistencies concerning their coverage and interoperability. An ecosystem, namely ComPath, has performed assessment of gene coverage

between KEGG, Reactome and WikiPathways [138] revealing an overlap of just 3800 human genes in the databases [139]. Additionally, the number of genes represented in similar pathways also varied largely. In this regard, Mubeen et al. (2019) have explored and illustrated the differences in the output of enrichment analyses influenced by choice of pathway databases. To resolve the issues of information discrepancies and coverage, the authors have merged several pathway databases to create an integrative database named as Mpath. The authors report on consistent and highly plausible results using Mpath as compared to results from individual pathway databases [140].

The causality (i.e., increment or decrement) in interacting biological entities is an important feature of biological networks, because it provides a clue about the direction of information flow and helps to distinguish upstream and downstream events. And causality, indeed, is the key ingredient for depicting chains of causal events, as it enables transformation of information from molecular levels to levels of mechanism. The interpretation of omics data through such causal biological networks (also known as HYPs) is called causal reasoning and bears the potential to elucidate insights which seemed impractical a few decades ago. A study by Chindelevitch et al. (2012) devised a scoring function to prioritize a large number of competing HYPs to identify the upstream causes of changes observed in gene expression profiles [141]. The authors were able to correctly identify upregulated oncogenes influenced by altered expression of upstream genes. In another similar study, individual HYPs representing RB1, MYC, TFRC and FOXO1 activities were identified as regulating signaling of AKT1, which is involved in cell proliferation and apoptosis [142].

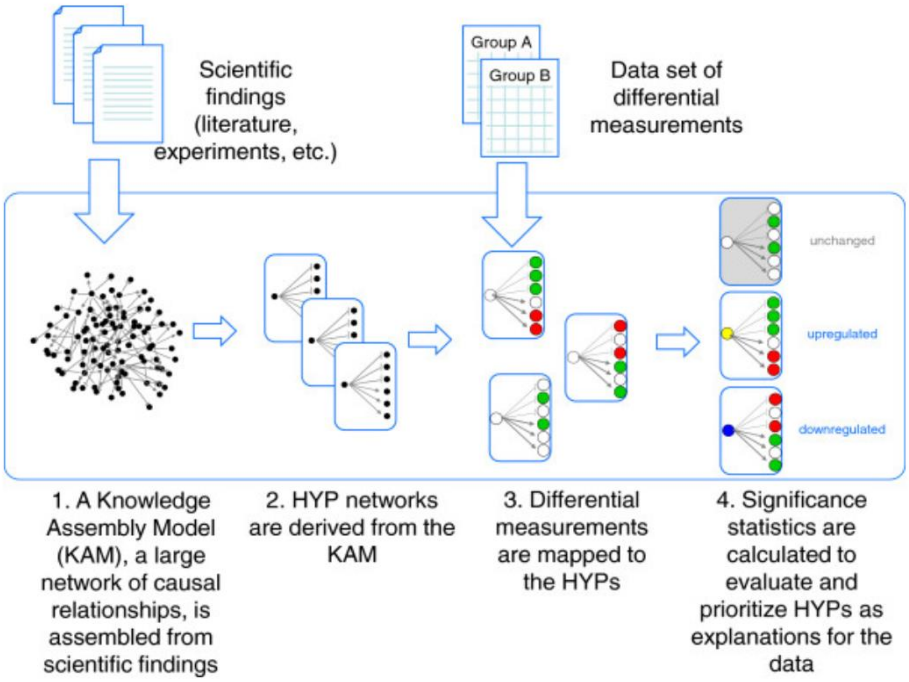

Figure 6: A workflow representing steps involved in the RCR algorithm where knowledge assembly derived molecular HYPs are mapped with omics data to identify upstream regulators that influence the expression of downstream entities. (Source: Catlett et al. (2013) [143])

Another notable algorithm known as Reverse Causal Reasoning (RCR), developed by Catlett et al. (2013), assesses the consistency of gene expression profiles with genes represented in HYPs by calculation of concordance and richness scores [143]. The algorithm identified potential molecular upstream regulators of observed differential gene expression in three different experimental settings. A general strategy of the RCR workflow is illustrated in Figure 6.

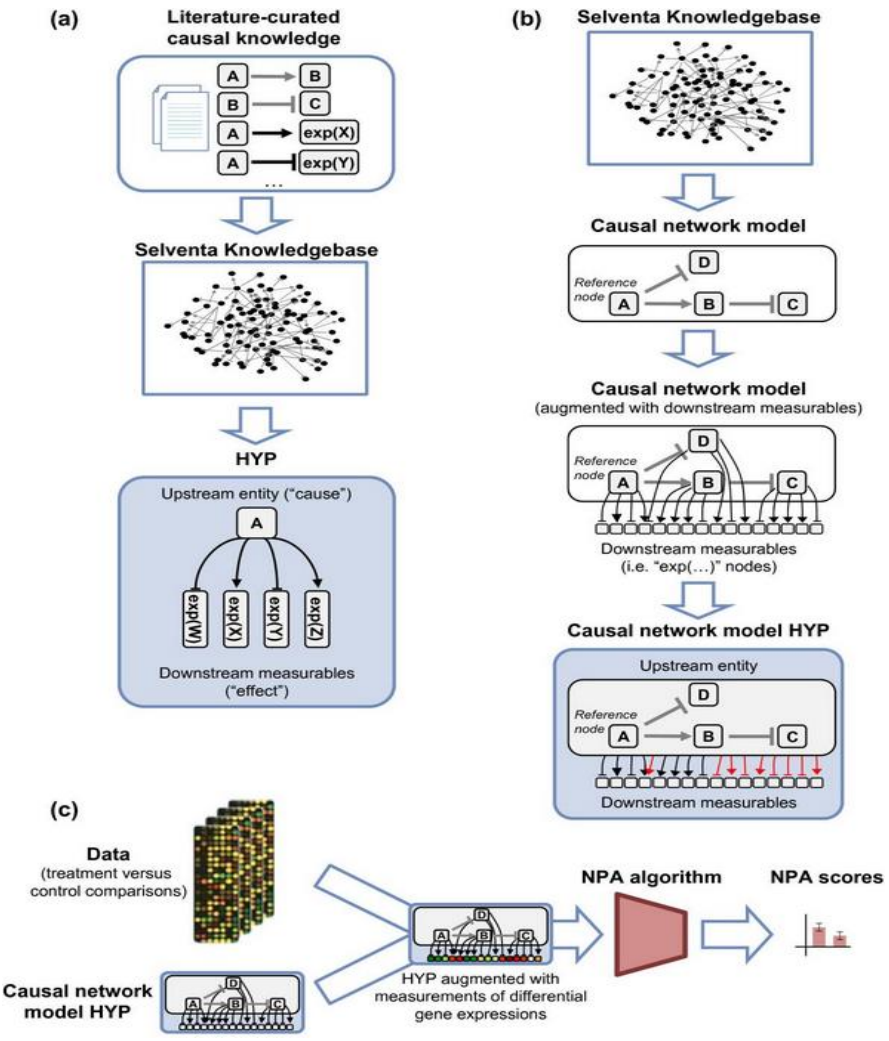

Figure 7: A schematic illustration of the NPA algorithm where knowledge assembly derived molecular HYPs representing molecular mechanisms are mapped with omics data and subjected to a scoring function that enables identification of most perturbed HYPs in treatment versus control conditions. (Source: Martin et al. (2012) [144])

Lastly, treatment-induced activity changes were quantified using the Network Perturbation Amplitude (NPA) algorithm, where causality of a molecular HYP is combined with gene expression profile to devise a scoring function [144]. The intensity of perturbation of a HYP as indicated by the NPA score allowed quantification of TNF-induced inflammatory signaling, suggesting that HYPs representing molecular mechanisms can be quantified. Figure 7 illustrates a general workflow of the NPA algorithm.

4. Thesis Outline

This thesis consists of three studies that have used systems biology approaches to address scientific issues in the field of neurodegenerative diseases (NDDs). Each of the studies presented here forms a bridge between the knowledge and data world of NDDs. Chapter 2 demonstrates a broader view about the comorbidity between AD and T2DM. By systematically encoding relevant biological information in formal statements, we showcase cross-talk between signaling pathways that are crucial for understanding AD-T2DM comorbidity, as their cross-talk gives rise to characteristic features of both diseases. Moreover, in this chapter, we use our mechanistic rationale to investigate the potential of repurposing an FDA-approved T2DM drug, namely, Metformin. In Chapter 3, we present a workflow that brings in the perspective of genetic variants in AD-T2DM comorbidity. The study is a purely data-driven approach, in which we perform functional assessment of important signals identified from analyses of genetic variants and gene expression profiles using knowledge assemblies. In Chapter 4, we discuss an algorithm that is able to quantify mechanistic graphs extracted from knowledge assemblies using gene expression profiles. We present two use cases, implementing the applying the algorithm to two mechanisms, one each for AD and PD. Finally, Chapter 5 summarizes the thesis by discussing limitations, outlook and future prospect of this thesis.

Chapter 2

Comorbidity Analysis between Alzheimer's Disease and Type 2 Diabetes Mellitus (T2DM) Based on Shared Pathways and the Role of T2DM Drugs

Introduction

The findings from a number of population-based, clinical and pathophysiology studies imply a comorbid association between AD and T2DM. From a chronological context, since the onset of T2DM takes place much earlier than the onset of AD, T2DM is considered a major risk factor for developing AD. Moreover, since the brains of AD patients are known to exhibit typical T2DM behavior (i.e., impaired insulin signaling), AD is sometimes referred to as T3DM. This is backed by the hypothesis that T2DM, a metabolic disorder in the pancreas, eventually matures in the brain as AD [56]. The comorbidity between AD and T2DM brings about some interesting and important inferences such as 1) Prevalence of shared molecular processes between the diseases 2) Possibility of designing similar treatment strategies 3) Prolonged use of medication for T2DM induces AD. In this work, using a systems biology modeling language, Biological Expression Language (BEL), mechanistic insights at the molecular level of AD and T2DM are deduced by addressing the aforementioned inferences. Furthermore, the modeling approach generates a broader view of what is already known about AD and T2DM, by incorporating discrete biological information and interactions mentioned in the literature. Lastly, the validation of the reconstructed mechanistic model was done through available omics data, namely, gene expression profiles, by identifying concordances and contradictions between formalized knowledge from the model and patterns in data.

Journal of Alzheimer's Disease 60 (2017) 721–731
DOI 10.3233/JAD-170440
IOS Press

Comorbidity Analysis between Alzheimer's Disease and Type 2 Diabetes Mellitus (T2DM) Based on Shared Pathways and the Role of T2DM Drugs

Reagon Karki[a,b], Alpha Tom Kodamullil[a,b] and Martin Hofmann-Apitius[a,b,*]
[a]*Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, Sankt Augustin, Germany*
[b]*Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn-Aachen International Center for IT, Bonn, Germany*

Accepted 20 July 2017

Abstract.
Background: Various studies suggest a comorbid association between Alzheimer's disease (AD) and type 2 diabetes mellitus (T2DM) indicating that there could be shared underlying pathophysiological mechanisms.
Objective: This study aims to systematically model relevant knowledge at the molecular level to find a mechanistic rationale explaining the existing comorbid association between AD and T2DM.
Method: We have used a knowledge-based modeling approach to build two network models for AD and T2DM using Biological Expression Language (BEL), which is capable of capturing and representing causal and correlative relationships at both molecular and clinical levels from various knowledge resources.
Results: Using comparative analysis, we have identified several putative "shared pathways". We demonstrate, at a mechanistic level, how the insulin signaling pathway is related to other significant AD pathways such as the neurotrophin signaling pathway, PI3K/AKT signaling, MTOR signaling, and MAPK signaling and how these pathways do cross-talk with each other both in AD and T2DM. In addition, we present a mechanistic hypothesis that explains both favorable and adverse effects of the anti-diabetic drug metformin in AD.
Conclusion: The two computable models introduced here provide a powerful framework to identify plausible mechanistic links shared between AD and T2DM and thereby identify targeted pathways for new therapeutics. Our approach can also be used to provide mechanistic answers to the question of why some T2DM treatments seem to increase the risk of AD.

Keywords: Alzheimer's disease, comorbidity, disease mechanisms, disease modeling, metformin, OpenBEL, type 2 diabetes mellitus

INTRODUCTION

Alzheimer's disease (AD) and type 2 diabetes mellitus (T2DM) are prevalent in aging populations. In particular, AD confronts us with the challenge of finding early stage diagnostic biomarkers that can be used for prevention and treatment and may help control the progression of the disease [1, 2]. In contrast, several classes of Food and Drug Administration (FDA) approved drugs like thiazolidinediones [3], DPP4-inhibitors [4], and GLP1 receptor agonists [5] are available for the treatment of T2DM. Despite the fact that T2DM is a metabolic disorder and AD is a central

*Correspondence to: Prof. Dr. Martin Hofmann-Apitius, Head of the Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), 53754 Sankt Augustin, Germany. Tel.: +49 2241 14 2802; Fax: +49 2241 14 2656; E-mail: martin.hofmann-apitius@scai.fraunhofer.de.

ISSN 1387-2877/17/$35.00 © 2017 – IOS Press and the authors. All rights reserved
This article is published online with Open Access and distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0).

nervous system disease, an increasing number of epidemiological studies suggest that there is a significant comorbid association between T2DM and AD [6, 7].

The comorbid association could be due to shared pathophysiology processes that are underlying both diseases [8, 9]. When we try to understand the putative shared pathophysiology of AD and T2DM, some important questions need to be addressed. Firstly, are the current methodologies capable and competent to provide better understanding of co-morbidity and their underlying pathophysiology processes? Secondly, how can we establish a mechanistic link between two medical conditions confined to specific regions of the body (brain for AD and liver for T2DM)? One of the major limitations of the conventional comorbidity measurement approaches is that they rely on clinical readouts and the association between these readouts is purely statistical. In order to establish mechanistic links between comorbid diseases, we systematically harvested and modeled relevant information on molecular mechanisms potentially shared by these diseases.

Another notion supporting the concept of shared mechanisms underlying comorbidity is based on the observation that the medication used for one disease could itself be a risk factor for the initiation or progression of another disease. There are many case reports about the initiation or modulation of a disease due to the usage of drugs for another indication. For example, drug-induced-parkinsonism has been observed in older patients due to the use of antipsychotic drugs such as haloperidol (HALDOL), chlorpromazine (THORAZINE) [10], thioridazine (MELLARIL) [11], trifluoperazine (STELAZINE) [12], and fluphenazine (PROLIXIN) [13]. The risk associated with antipsychotics is often dose dependent and related to dopamine D2 striatal occupancy, which is linked mechanistically to parkinsonism [14, 15].

Previous attempts to understand the comorbid association between AD and T2DM have not considered the role of dysregulated entities such as genes/proteins, SNPs, and miRNAs and impaired biological processes involved in the diseases but rather have focused on specific biological pathways of interest. Associations between biological pathways and comorbid observations are usually reported in the form of free text; whereas pathway information is commonly represented in various pathway databases. A context-specific, knowledge-based network modeling approach, however, may provide a better way to integrate all the scientific knowledge around comorbid diseases and to identify common underlying mechanisms. Motivated by the capabilities of the Biological Expression Language (BEL) [16] to construct cause and effect computable network models, we have therefore generated AD and T2DM models based on knowledge extracted from the scientific literature. The resulting mechanistic network modeling work was driven by two hypotheses: 1) Impaired pathways in T2DM increase the risk for AD, and 2) T2DM drugs increase the risk of AD. Using our mechanistic modeling approach, we have tried to unravel shared pathways possibly perturbed by a drug prescribed for one of the comorbid diseases which could be causally involved in the etiology the other comorbid disease. The models developed here not only represent a comprehensive view on shared pathways between the two diseases but also provide a means to mechanistically differentiate the effects induced by treatments and explain how it contributes to comorbidity.

METHODS

Data collection and model building

Firstly, SCAIView [17], a tool that allows semantic search and retrieval of articles, was used to build well-defined literature corpora. Secondly, all the articles were manually checked for their relevancy to the context of diseases. Thirdly, BEL coding experts read through articles to extract essential lines of evidence; which were subsequently encoded into BEL statements. We have considered a total of 448 articles, which were manually converted to BEL statements to build the AD model. Similarly, a total of 106 articles that explicitly focus on the shared mechanisms of T2DM with AD were considered to build the T2DM model. We have also extracted relevant information from databases after manually checking the referenced articles. In this way, two comprehensive systems biology models specific to AD and T2DM were built from mostly human-based PubMed articles and available pathway databases like KEGG and Reactome.

Identification of common pathways between AD and T2DM

To identify shared pathways, which are enriched in our models of AD and T2DM, we have performed gene set enrichment analysis using a functional annotation tool provided by the Database for Annotation,

Visualization and Integrated Discovery (DAVID) [18]. The tool enables users to retrieve a wide range of annotations, mainly GO terms, biological pathways, protein-protein interactions, and disease associations, represented by a given list of genes. The significance of each resulting annotation is determined by a *p*-value calculated by modified Fisher Exact test, where the smaller the *p*-value, the more significant is the output. We deployed the tool to functionally annotate the list of genes from our models with biological processes or pathways. Sub-graphs representing canonical pathways were extracted from the models and analyzed to identify shared edges and nodes amongst the enriched pathways.

Investigation of the role of T2DM drugs to AD pathology

Using SCAIView, we retrieved all drugs mentioned in the literature for T2DM with the idea in mind to analyze, whether T2DM drugs have been reported to cause AD. We found 1,060 entries/drug names from 78,248 documents, which were ranked based on relative entropy [19]. We further analyzed the role of top 20 T2DM drugs based on the mechanism derived from our AD models. The aim of this analysis was to identify T2DM treatments that could be associated with the development of AD.

RESULTS

Causal and correlative BEL models representing mechanistic pathways in AD and T2DM

Using a literature mining approach, we selected 448 and 106 articles, which were found to contain relevant information about AD and T2DM, respectively. The AD BEL model consists of 2,004 nodes and 4,766 edges representing 3,068 BEL statements. The nodes consist of 539 proteins, 273 biological processes, 176 SNPs, 163 complexes, 140 chemical entities, 136 genes, 45 RNAs, 41 miRNAs, 23 pathologies, and 468 other entities representing translocation, transcription, and degradation processes. Similarly, in the context of T2DM, we have extracted 1,333 BEL statements to build a network comprising of 1,094 nodes and 2,414 edges. The nodes consist of 183 proteins, 146 biological processes, 327 SNPs, 48 complexes, 86 chemical entities, 92 genes, 24 RNAs, 12 miRNAs, 28 pathologies, and 148 other entities representing processes like translocation, transcription, and degradation. There are about 31 commonly impaired bio-processes (mainly: insulin resistance, insulin signaling pathway, oxidative stress, mitochondrial dysfunction, degradation of beta cells, neuron apoptosis, etc.), which are associated with both AD and T2DM. Likewise, 9 common diseases/pathologies like cardiovascular disorders, obesity, and amyloidosis were found to be common to both AD and T2DM models, which are often co-mentioned with AD and T2DM.

Cross talk between insulin signaling pathway and other AD specific pathways with respect to AD and T2DM

In order to identify shared signaling pathways perturbed in both AD and T2DM, we have performed systematic comparisons of the two models based on gene sets (pathways) derived from the models by applying the DAVID tool. This analysis allowed us to prioritize the pathways already enriched in our models and to further extract shared mechanisms or sub-networks common to both models. Among others, we have identified insulin signaling pathway, neurotrophin signaling pathway, PI3K/AKT signaling, MTOR signaling, MAPK signaling, and microglial mediated immune responses as the top-ranked pathways. We analyzed further how these specific pathways do cross-talk to each other, potentially contributing to the comorbidity between AD and T2DM.

As shown in Fig. 1, in normal insulin signaling pathway, insulin (INS) binds to the insulin receptor (INSR) causing a phosphorylation of INSR, thereby activating INSR to bind to insulin like growth factor 1 receptor (IGF1R). This interaction phosphorylates IGF1R to further activate insulin receptor substrate 2 (IRS2) and insulin receptor substrate 4 (IRS4). In AD, IRS2 and IRS4 interact with protein tyrosine phosphatase, non-receptor type 11 (PTPN11) and phosphoinositide-3-kinase regulatory subunit 1 (PIK3R1) to activate phosphatidylinositol-4, 5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA) which increase phosphatidylinositol (3, 4, 5)-trisphosphate (PIP3). Similar to these events in AD, PIP3 activation by insulin receptor substrates through PIK3CA and their receptors has been observed in T2DM. PIP3 increases apoptosis by increased phosphorylation of BCL2 associated agonist of cell death (BAD) and activation of forkhead box O3 (FOXO3) through AKT serine/threonine kinase 1 (AKT1) [20]. Furthermore, AKT1 hyper-activates the mechanistic target of

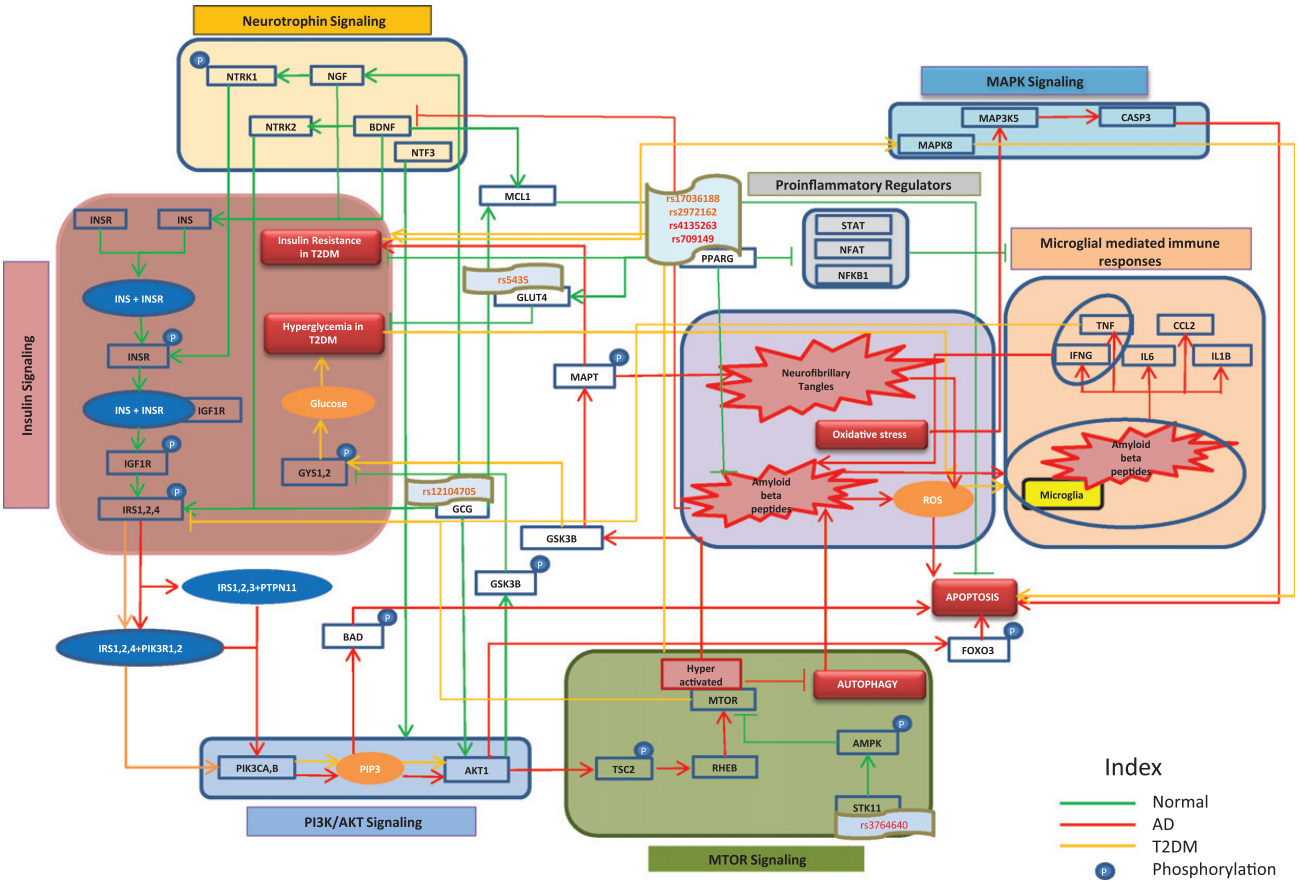

Fig. 1. Cross talk between significant pathways in AD and T2DM. The cartoon represents interactions among entities of different signaling pathways involved in AD and T2DM. Red, orange, and green edges represent AD, T2DM, and normal conditions, respectively. Here, we depict the role of insulin signaling pathway and involvement of other pathways like PI3K/AKT signaling, MOTR signaling, MAPK signaling, and Neurotrophin signaling in AD and T2DM which give rise to characteristic features of both the disease.

rapamycin (MTOR) through increased activation of a Ras homologue enriched in brain (RHEB) by phosphorylating the tuberous sclerosis 2 protein (TSC2) [21]. Inactivation of MTOR promotes autophagy, thereby regulating the process of removal of amyloid-β (Aβ) from the brain [22, 23]. The presence of Aβ hinders the activity of brain derived neurotrophic factor (BDNF), which is known to regulate INS [24]. In T2DM, hyperactivated MTOR inhibits normal phosphorylation of insulin substrates, affecting insulin sensitivity and thereby increasing insulin resistance, which in turn leads to apoptosis through upregulation of mitogen-activated protein kinase 8 (MAPK8) [25]. Likewise, tumor necrosis factor (TNF) in T2DM is also responsible for abnormal serine phosphorylation of insulin substrates [26]. In normal brain, AKT1 is involved in phosphorylating glycogen synthase kinase 3β (GSK3β), restricting the ability of GSK3β to further phosphorylate microtubule associated protein tau (MAPT), which will otherwise lead to increased deposition of that protein in neurofibrillary tangles (NFTs) [27]. Moreover, it also regulates the intake of glucose by decreasing phosphorylation of glycogen synthase 1 (GYS1) and glycogen synthase 2 (GYS2) [28, 29]. We found that AKT1 is consistently downregulated in AD as well as in T2DM, which supports the fact that glucose levels are increased in blood, a condition called hyperglycemia [30]. Hyperglycemia in T2DM increases reactive oxygen species, thereby becoming detrimental to normal beta cell function. Similarly in AD, NFTs and Aβ increase oxidative stress by producing reactive oxygen species, leading to activation of microglia and inflammatory regulators [31, 32]. The increase in oxidative stress reported in AD, increases levels of mitogen-activated protein kinase kinase kinase 5 (MAP3K5), which further adds to caspase 3 (CASP3) activity resulting in apoptosis [33].

To support the above-mentioned pathways and mechanisms, we analyzed gene expression datasets

(Supplementary File 1 and 2) to understand the expression patterns of the entities involved in AD and T2DM. The expressions of protein kinase AMP-activated catalytic subunits (PRKAA1 and PRKAA2) and protein kinase AMP-activated non-catalytic subunits (PRKAB2 and PRKAG3), which code for AMP-activated protein kinase (AMPK) sub-units, were found to be downregulated in AD, which is consistent with the above depicted Fig. 1. Similarly, protein kinase AMP-activated non-catalytic subunit gamma 1 (PRKAG1) and protein kinase AMP-activated non-catalytic subunit gamma 2 (PRKAG2), which also code for AMPK sub-units, were found to be downregulated in T2DM. Moreover, serine/threonine kinase 11 (STK11), a protein that activates AMPK [34], was found to be underexpressed in both AD and T2DM. Likewise, we found downregulated signal transducer and activator of transcription (STAT1 and STAT2), nuclear factors of activated T-cells (NFATC1, NFATC2, and NFATC3), and nuclear factor kappa B subunit 1 (NFKB1) to be associated with AD conditions, all of which are pro-inflammatory regulators associated with microglia-mediated immune response. Correspondingly, in the context of T2DM, downregulation of STAT1 and nuclear factor of activated T-cells 4 (NFATC4) were observed. The expressions of neurotrophic receptor tyrosine kinases (NTRK1 and NTRK2) and neurotrophin 3 (NTF3), which regulate insulin-signaling pathway, were found to be down regulated in AD, while only NTRK2 was downregulated in T2DM. However, we did not find any significant signals regarding the expression of glucagon (GCG), solute carrier family 2 member 4 (SLC2A4), peroxisome proliferator activated receptor gamma (PPARG), and BCL2 family apoptosis regulator (MCL1) from the data sets we analyzed. Nevertheless, we could identify certain SNPs that are associated with T2DM and AD in patient cohorts through literature. We identified SNP rs12104705 to be associated with GCG [35], SNP rs5435 with SLC2A4 [36], and SNPs rs1801282 and rs1805192 with PPARG [37, 38]. These genetic variants may contribute to the perturbation of normal functions of these genes/proteins in the disease state. A dedicated mechanistic analysis and more independent cohort studies are needed to prove the functional role of these genetic variants in causing the comorbidity. One such analysis identifying STK11 genetic variant rs3764640 in regulating autophagy has been depicted by Kodamullil et al. [39].

Comorbidity analysis based on use of drugs

Using our in-house text mining tool, SCAIView, we retrieved FDA-approved drugs from AD and T2DM to understand the perspective of comorbidity in the context of drugs. Among the 5 approved AD drugs, we found that only tacrine has some effects on T2DM [40, 41]. In contrast, 20 approved T2DM drugs have already been investigated for repurposing for AD (Supplementary File 3) targeting mainly PPARG, AMPK, GCG, and leptin (LEP). In the following sections, we briefly discuss positive and negative effects of various drug targets taking into account relevant studies (mostly human based and few animal based) and further elaborate on the effects of metformin in AD.

PPARG functions by regulating SLC2A4, a protein that plays a vital role in T2DM as it enhances transportation and absorption of glucose [42]. Since a type of brain-specific-diabetes is observed in AD [43], targeting PPARG in AD has been widely considered. Moreover, PPARG is capable of inhibiting pro-inflammatory regulators, which are responsible for microglial activation [44]. The over-activation of microglia has been shown to lead to deposition of Aβ peptides through excess release of inflammatory factors in AD [45]. The second common target, AMPK, improves glucose metabolism and insulin sensitivity in T2DM, a much-needed activity in the normal brain [46, 47]. For this reason, T2DM drugs targeting AMPK are considered as repurposing candidates for AD. The other common targets, GCG and LEP, are also interesting targets in T2DM. Although they are not directly involved in the insulin signaling pathway, they can have an effect on AD pathophysiology mechanisms. Inducing glucose lowering effects in T2DM, GCG is able to improve synaptogenesis and neurogenesis, inhibit depositions of Aβ and microglial activation in AD [48]. Likewise, LEP activation has been shown to reduce enzymatic activity of beta-secretase 1 (BACE1) and phosphorylation of MAPT [49].

However, some contradictions reporting the findings that T2DM drugs may increase the risk of AD keep the chances of repurposing T2DM drugs at bay. Rosiglitazone, an agonist of PPARG, has a very low blood-brain barrier penetration [50]. Thus, induced insulin sensitivity in the brain might not be good enough to regulate normal insulin signaling. On the other hand, as it effectively sensitizes peripheral tissues to insulin, the levels of blood insulin are remarkably decreased, thereby reducing insulin

levels in brain. This is assumed to promote neuronal insulin resistance over time [51]. Metformin, a T2DM drug, is known to upregulate expression of BACE1 and increase deposition of Aβ [52]. Furthermore, sitagliptin has been reported to increase MAPT phosphorylation, eventually leading to NFTs [53]. In this regard, it can be concluded that T2DM drugs amplify the risk to develop AD or at least serve as a risk factor in AD. This highlights the need of shifting research works from repurposing T2DM drugs in AD to not-well-known possible influences of T2DM drugs in developing AD. Above all, it is of utmost importance to understand the mechanisms modified by T2DM drugs that will eventually lead to AD.

Identification of the role of metformin in AD using disease models

To identify the potential role of drugs in causing or at least modulating the risk of comorbidity between diseases, we have performed a mechanistic analysis of the top T2DM drug targets in the context of AD. The search of drugs related to T2DM using SCAIView indicated that metformin belongs to the drugs with highest relevance (based on relative entropy scores) [54]. Metformin is a FDA-approved biguanide anti-hyperglycemic agent used for the treatment of T2DM. The mechanism of action of metformin is understood to reduce blood glucose levels by decreasing glucose production in the liver, reducing intestinal absorption of glucose and improving insulin sensitivity by increasing uptake and utilization of glucose in the peripheral regions of the body [55]. Modeling the drug-target-pathway context of metformin resulted in a complex pattern: we were able to explain, at a mechanistic level, the discrepancy of epidemiological observations that are linked to effects of drug treatment. These mechanistic explanations are in sharp contradiction to previously suggested opportunities of repurposing metformin in AD. We can, however, also reconstruct the most likely mechanistic explanation for the beneficial effects of metformin. It remains to be shown, how far genetic variation (SNPs) and epigenetics effects account for the differences observed in epidemiological studies.

Putative beneficial effects

Since AD is frequently accompanied by insulin resistance [56, 57], metformin is sought to improve insulin sensitivity in AD patients. It is also capable of inhibiting MAPT phosphorylation through increased activity of PP2A as evident from a study from Kickstein et al. [58]. The authors explain this mechanism by reporting the finding that metformin in fact interrupts the binding of PP2A with MID1/α4 complex, a protein-protein interaction involved in degradation of PP2A. Hence, it is likely to prevent formation of NFTs and reduce progression of AD [58]. Similarly, it inhibits neuronal damage via upregulation of glucagon like peptide 1 receptor (GLP1R), a glucagon receptor [59]. The combined use of metformin and INS is reported to reduce the aggregation of Aβ [52]. Furthermore, the drug is known to inhibit the JNK cascade, formation of advanced glycation end products (AGE) and protect against degradation of synaptophysin (SYP), a protein involved in synaptic transmission [60, 61]. In addition, AMPK's ability to promote cell survival is surged by metformin [62]. Through our gene expression analysis of mice samples (Supplementary File 1), we observed that PP2As and AMPKs increased with metformin treatment while SYP was found to be decreased. A simple cartoon representation of beneficial effects of metformin is shown below in Fig. 2.

Putative harmful effects

The following "chain of causation" (Fig. 3) provides a mechanistic explanation for the observed harmful effects of metformin. We assume that "modifiers" (e.g., the genetic makeup of individual or epigenetic modifications) outside of our models may contribute to the overall decision making process that results in either beneficial or harmful effects.

Metformin adds to deposition of Aβ by increasing transcriptional activity of BACE1 [52]. Picone et al. [63] report the findings that it contributes in accumulation of Aβ through NFKB1 activation which further upregulates PSEN1 and APP. The same study reveals that treatment with metformin increased oxidative stress and mitochondrial damage and decreased expressions of Cytochrome C (CYCS) and Hexokinase 2 (HK2). As a result of these effects of metformin, cell death was observed [63]. In addition, it promotes insoluble tau aggregation as reported by Barini et al. suggesting that it could possibly increase the risk of tauopathy among metformin treated diabetic patients [64]. The drug also reduces activity of vitamin B12 which causes reduced epidermal growth factor (EGF) and increased TNF [65], the latter of which is often associated with apoptosis and neuroinflammation. In normal conditions, EGF has a positive effect on nervous system development

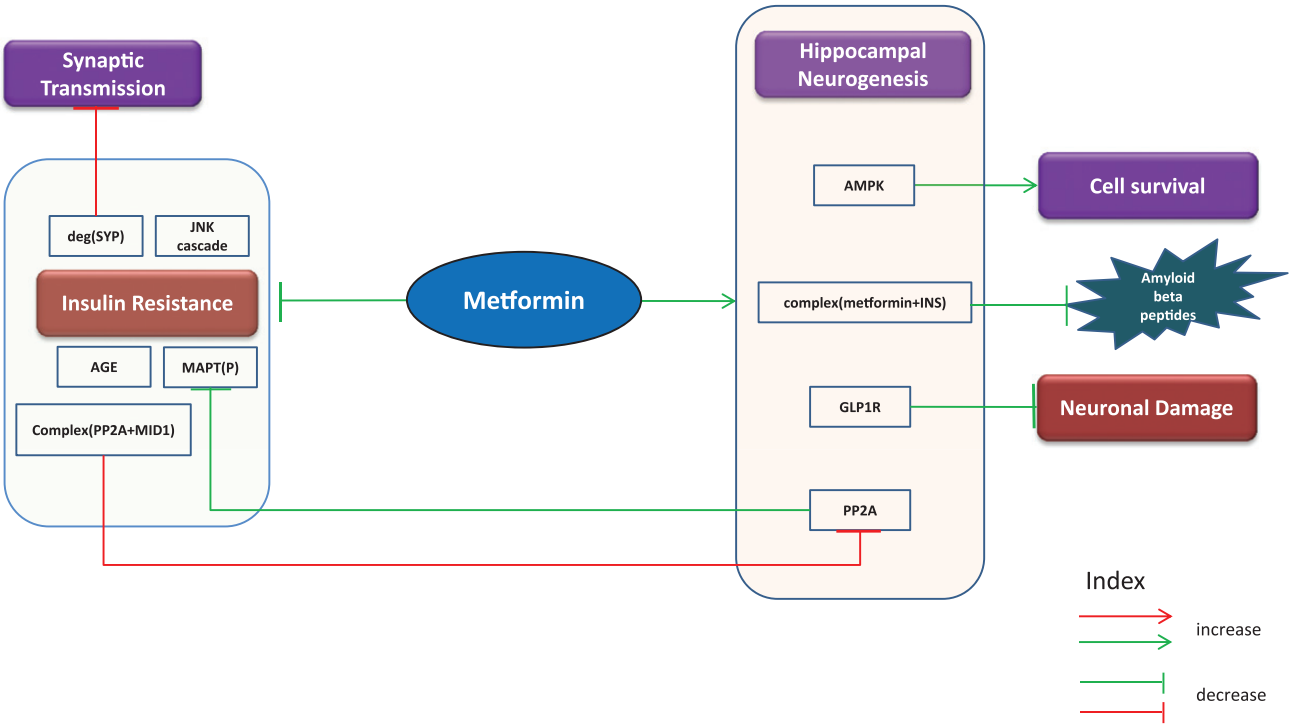

Fig. 2. Putative beneficial effects. Putative beneficial effects of metformin represented as cartoon diagram: The green edges refer to effects induced by metformin or events in normal conditions while the red edges indicate events in diseased state. The capability of metformin to reduce insulin resistance and neuronal damage, promote cell survival and hippocampal neurogenesis and inhibit AGEs, JNK cascade, and phosphorylation of MAPT provides us with the opportunity to repurpose metformin for AD.

and levels of A Disintegrin And Metalloproteinase domain-containing protein 10 (ADAM10), a protein known to reduce Aβ aggregation [66]. The expression patterns of the aforementioned genes (Supplementary File 1) were analyzed to understand their concordances with literature. The upregulated expressions of BACE1, PSEN1, and APP identified from gene expression analyses of mice and human samples are consistent with the literature findings. In contrast, expression of EGF in mice was observed to be downregulated.

DISCUSSION

A large number of epidemiological, preclinical, and pathophysiology studies indicate that AD and T2DM share cellular and molecular mechanisms. The classical approaches in measuring comorbidity that are based on clinical readouts, patient data, and electronic health records cannot reason over the dysfunctional molecular activity or the impaired biological pathway involved in the diseased state. On the contrary, deciphering comorbidity at a mechanistic level could well explain the outcomes of clinical readouts and patient examinations establishing a link between proteomic/genomic and phenotypic aspects of diseases. However, in this study we do not attempt to cover this proposal. Since there are no established studies aimed at explaining comorbidity based on shared mechanisms, we believe that understanding the co-morbid mechanisms between complex diseases can be dealt with systems biology approaches like integrative modeling. We followed a knowledge-driven modeling approach, which served as a rationale to infer the mechanistic background of comorbidity association between AD and T2DM. Modeling using BEL bears a high "explanatory" potential; although we do not necessarily discover new knowledge, we bring information into context and are able to reconstruct mechanisms. As we are aware of the publication bias, we have therefore done model-validation through available data (gene expression profiles) as the key to identify contradictions or concordances between formalized knowledge and patterns in data.

The synopsis of mechanisms relevant for T2DM and AD reveals that there is crosstalk among important pathways that play either a role in T2DM or AD and are thus candidates for shared pathways possibly involved in the observed comorbidity. The results of this study demonstrate that encoding rel-

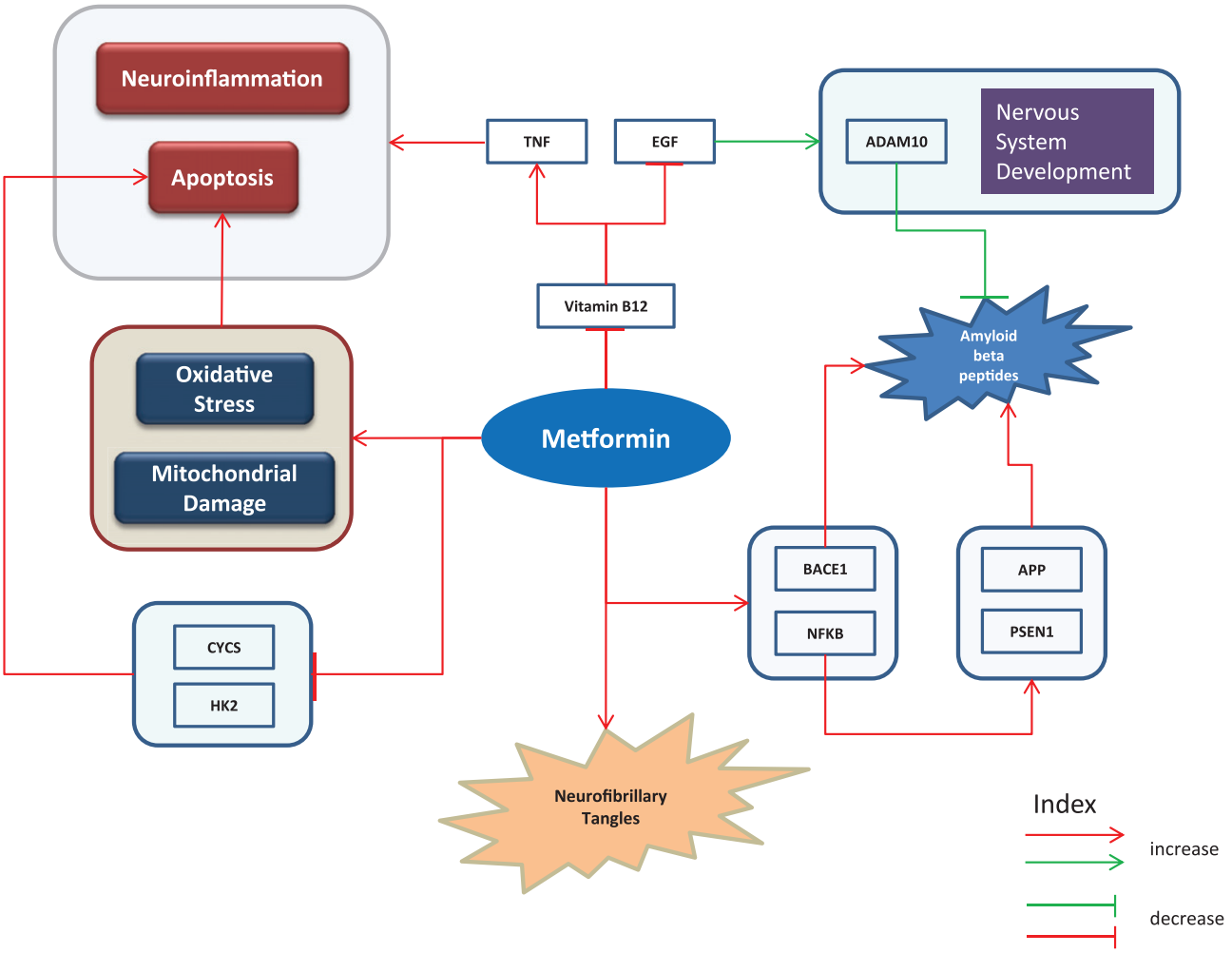

Fig. 3. Putative harmful effects. Harmful effects of metformin represented as cartoon diagram: The red edges represent effects of metformin or events observed in diseased state while the green edges refer to events in normal conditions. Metformin's use has been reported to contribute to characteristic features of AD such as apoptosis, neuroinflammation, neurofibrillary tangle formation, and aggregated amyloid-β questioning the suggested opportunities seen in metformin as a repurposible drug.

evant knowledge into causal relationship models confers enhanced interpretation power that is well-suited for comorbidity analysis. Our results provide additional support to previously suggested comorbid associations between AD and T2DM. We have shown at a mechanistic level how entities involved in insulin signaling, PI3K/AKT signaling, MTOR signaling, neurotrophin signaling, and microglial-mediated immune responses interact and potentially contribute to the manifestation of characteristic features of both AD and T2DM. Since there are a few inconsistencies in the gene expression data (both humans and mice) to support the key interactions depicted in this paper, integration of genetic variants into the models may add to the explanatory power of the models and support the notion of candidate comorbid mechanisms. Furthermore, depicting the AD features induced by metformin, we hypothesize that drug treatment itself could contribute to the comorbidity between AD and T2DM. A study aimed at understanding the progression of AD by comparing metformin-treated-T2DM patients with other T2DM patients treated with other drugs is needed to validate the effect. This emphasizes the need to reconsider the prescription of drugs if there is any evidence of comorbid disease associated with any drug.

It is clear from this work that BEL based network modeling approaches bear great potential to help us to identify shared mechanisms between two diseases. However, as new knowledge is being generated and communicated all the time, we need a continuous update of the models in order to unravel new mechanisms and to take into account additional factors that may contribute to comorbidity. Additionally, as complex diseases like AD as well as T2DM progress with time, there is also a need to integrate the time

dependent cascade of events, which is not currently dealt with by BEL modeling. Our hope is that this sort of analysis will allow us to identify new drug repurposing candidates based on the common mechanism between diseases. Based on our analysis, we see a need to re-evaluate the role of existing drugs because besides having positive effects against a particular disease, they might also be involved in progression of another disease. What needs to be addressed in the future is the definition of the role of genetic variants and epigenetic modifications in order to generate a comprehensive picture of the mechanisms underlying comorbidity of diseases together with the time dependences. Given the identification of candidate mechanisms for comorbidity between AD and T2DM, we propose additional experiments around these pathways to find common targets between AD and T2DM, which could pave way for new therapeutic developments that take shared mechanisms into account.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from the B-IT foundation that sponsors part of the academic work in our department. Part of the research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115568 (project AETIONOMY), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2017-2013) and European Federation of Pharmaceutical Industries and Associations companies' in kind contribution.

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/17-0440r1).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/jad-170440.

REFERENCES

[1] Wang KC, Woung LC, Tsai MT, Liu CC, Su YH, Li CY (2012) Risk of Alzheimer's disease in relation to diabetes: A population-based cohort study. *Neuroepidemiology* **38**, 237-244.

[2] Huang CC, Chung CM, Leu HB, Lin LY, Chiu CC, Hsu CY, Chiang CH, Huang PH, Chen TJ, Lin SJ, Chen JW, Chan WL (2014) Diabetes mellitus and the risk of Alzheimer's disease: A nationwide population-based study. *PLoS One* **9**, e87095.

[3] Richter B, Bandeira-Echtler E, Bergerhoff K, Clar C, Ebrahim SH (2007) Rosiglitazone for type 2 diabetes mellitus. *Cochrane Database Syst Rev*, CD006063.

[4] Lee M, Rhee MK (2015) Sitagliptin for type 2 diabetes: A 2015 update. *Expert Rev Cardiovasc Ther* **13**, 597-610.

[5] Kalra S (2013) Glucagon-like peptide-1 receptors agonists (GLP1 RA). *J Pak Med Assoc* **63**, 1312-1315.

[6] Yogi-Morren D, Galioto R, Strandjord SE, Kennedy L, Manroa P, Kirwan JP, Kashyap S, Gunstad J (2014) Duration of type 2 diabetes and very low density lipoprotein levels are associated with cognitive dysfunction in metabolic syndrome. *Cardiovasc Psychiatry Neurol* **2014**, 656341.

[7] Cha DS, Carvalho AF, Rosenblat JD, Ali MM, McIntyre RS (2014) Major depressive disorder and type II diabetes mellitus: Mechanisms underlying risk for Alzheimer's disease. *CNS Neurol Disord Drug Targets* **13**, 1740-1749.

[8] Shaik MM, Gan SH, Kamal MA (2014) Epigenomic approach in understanding Alzheimer's disease and type 2 diabetes mellitus. *CNS Neurol Disord Drug Targets* **13**, 283-289.

[9] Moreira PI (2012) Alzheimer's disease and diabetes: An integrative view of the role of mitochondria, oxidative stress, and insulin. *J Alzheimers Dis* **30**, S199-S215.

[10] Shin H-W, Chung SJ (2012) Drug-induced parkinsonism. *J Clin Neurol* **8**, 15-21.

[11] O'Keeffe R, Sharman DF, Vogt M (1970) Effect of drugs used in psychoses on cerebral dopamine metabolism. *Br J Pharmacol* **38**, 287-304.

[12] Bashford G, Bradd P (1996) Drug-induced Parkinsonism associated with dysphagia and aspiration: A brief report. *J Geriatr Psychiatry Neurol* **9**, 133-135.

[13] Dreyfuss J, Beer B, Devine DD, Roberts BF, Schreiber EC (1972) Fluphenazine-induced parkinsonism in the baboon: Pharmacological and metabolic studies. *Neuropharmacology* **11**, 223-230.

[14] Montastruc JL, Llau ME, Rascol O, Senard JM (1994) Drug-induced parkinsonism: A review. *Fundam Clin Pharmacol* **8**, 293-306.

[15] Bohlega SA, Al-Foghom NB (2013) Drug-induced Parkinson's disease. *Neurosciences (Riyadh)* **18**, 215-221.

[16] Biological Expression Language. http://www.openbel.org/. Accessed December 8, 2016.

[17] SCAIView. http://www.scaiview.com/. Accessed December 5, 2016.

[18] Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13.

[19] Younesi E, Toldo L, Müller B, Friedrich CM, Novac N, Scheer A, Hofmann-Apitius M, Fluck J (2012) Mining biomarker information in biomedical literature. *BMC Med Inform Decis Mak* **12**, 148.

[20] Freude S, Schilbach K, Schubert M (2009) The role of IGF-1 receptor and insulin receptor signaling for the pathogenesis of Alzheimer's disease: From model organisms to human disease. *Curr Alzheimer Res* **6**, 213-223.

[21] Dibble CC, Cantley LC (2015) Regulation of mTORC1 by PI3K signaling. *Trends Cell Biol* **25**, 545-555.

[22] Kim S, Choi KJ, Cho S-J, Yun S-M, Jeon J-P, Koh YH, Song J, Johnson GVW, Jo C (2016) Fisetin stimulates autophagic degradation of phosphorylated tau via the activation of TFEB and Nrf2 transcription factors. *Sci Rep* **6**, 24933.

[23] Caccamo A, Majumder S, Richardson A, Strong R, Oddo S (2010) Molecular interplay between mammalian target of rapamycin (mTOR), amyloid-β, and tau effects on cognitive impairments. *J Biol Chem* **285**, 13107-13120.

[24] Ye X, Tai W, Zhang D (2012) The early events of Alzheimer's disease pathology: From mitochondrial dysfunction to BDNF axonal transport deficits. *Neurobiol Aging* **33**, 1122.e1-10.

[25] Chai W, Liu Z (2007) p38 mitogen-activated protein kinase mediates palmitate-induced apoptosis but not inhibitor of nuclear factor-κB degradation in human coronary artery endothelial cells. *Endocrinology* **148**, 1622-1628.

[26] Gual P, Le Marchand-Brustel Y, Tanti J-F (2005) Positive and negative regulation of insulin signaling through IRS-1 phosphorylation. *Biochimie* **87**, 99-109.

[27] Kettunen P, Larsson S, Holmgren S, Olsson S, Minthon L, Zetterberg H, Blennow K, Nilsson S, Sjölander A (2015) Genetic variants of GSK3B are associated with biomarkers for Alzheimer's disease and cognitive function. *J Alzheimers Dis* **44**, 1313-1322.

[28] Liu Y, Liu F, Grundke-Iqbal I, Iqbal K, Gong CX (2011) Deficient brain insulin signalling pathway in Alzheimer's disease and diabetes. *J Pathol* **225**, 54-62.

[29] Bevan P (2001) Insulin signalling. *J Cell Sci* **114**, 1429-1430.

[30] Sridhar GR, Lakshmi G, Nagamani G (2015) Emerging links between type 2 diabetes and Alzheimer's disease. *World J Diabetes* **6**, 744-751.

[31] Schilling T, Eder C (2011) Amyloid-β-induced reactive oxygen species production and priming are differentially regulated by ion channels in microglia. *J Cell Physiol* **226**, 3295-3302.

[32] Dumont M, Beal MF (2011) Neuroprotective strategies involving ROS in Alzheimer disease. *Free Radic Biol Med* **51**, 1014-1026.

[33] Inoue H, Tsukita K, Iwasato T, Suzuki Y, Tomioka M, Tateno M, Nagao M, Kawata A, Saido TC, Miura M, Misawa H, Itohara S, Takahashi R (2003) The crucial role of caspase-9 in the disease progression of a transgenic ALS mouse model. *EMBO J* **22**, 6665-6674.

[34] Koh HJ (2016) Regulation of exercise-stimulated glucose uptake in skeletal muscle. *Ann Pediatr Endocrinol Metab* **21**, 61-65.

[35] Li L, Gao K, Zhao J, Feng T, Yin L, Wang J, Wang C, Li C, Wang Y, Wang Q, Zhai Y1, You H, Ren Y, Wang B, Hu D (2014) Glucagon gene polymorphism modifies the effects of smoking and physical activity on risk of type 2 diabetes mellitus in Han Chinese. *Gene* **534**, 352-355.

[36] Bodhini D, Radha V, Ghosh S, Majumder PP, Rao MRS, Mohan V (2011) GLUT4 gene polymorphisms and their association with type 2 diabetes in south Indians. *Diabetes Technol Ther* **13**, 913-920.

[37] Kasim NB, Huri HZ, Vethakkan SR, Ibrahim L, Abdullah BM (2016) Genetic polymorphisms associated with overweight and obesity in uncontrolled Type 2 diabetes mellitus. *Biomark Med* **10**, 403-415.

[38] Scacchi R, Pinto A, Gambina G, Rosano A, Corbo RM (2007) The peroxisome proliferator-activated receptor gamma (PPAR-γ2) Pro12Ala polymorphism is associated with higher risk for Alzheimer's disease in octogenarians. *Brain Res* **1139**, 1-5.

[39] Kodamullil AT, Younesi E, Naz M, Bagewadi S, Hofmann-Apitius M (2015) Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis. *Alzheimers Dement* **11**, 1329-1339.

[40] Zhao Q, Matsumoto K, Tsuneyama K, Tanaka K, Li F, Shibahara N, Miyata T, Yokozawa T (2011) Diabetes-induced central cholinergic neuronal loss and cognitive deficit are attenuated by tacrine and a Chinese herbal prescription, kangen-karyu: Elucidation in type 2 diabetes db/db mice. *J Pharmacol Sci* **117**, 230-242.

[41] Niu Y, Li F, Inada C, Tanaka K, Watanabe S, Fujiwara H, Sasaki-Hamada S, Oka JI, Matsumoto K (2015) Chemical profiling with HPLC-FTMS of exogenous and endogenous chemicals susceptible to the administration of chotosan in an animal model of type 2 diabetes-induced dementia. *J Pharm Biomed Anal* **104**, 21-30.

[42] Rangwala SM, Lazar MA (2004) Peroxisome proliferator-activated receptor γ in diabetes and metabolism. *Trends Pharmacol Sci* **25**, 331-336.

[43] Suzanne M, Wands JR (2008) Alzheimer's disease is type 3 diabetes-evidence reviewed. *J Diabetes Sci Technol* **2**, 1101-1113.

[44] Landreth GE, Heneka MT (2001) Anti-inflammatory actions of peroxisome proliferator-activated receptor gamma agonists in Alzheimer's disease. *Neurobiol Aging* **22**, 937-944.

[45] Xing H, Guo S, Zhang Y, Zheng Z, Wang H (2016) Upregulation of microRNA-206 enhances lipopolysaccharide-induced inflammation and release of amyloid-β by targeting insulin-like growth factor 1 in microglia. *Mol Med Rep* **14**, 1357-1364.

[46] Shi L, Zhang T, Zhou Y, Zeng X, Ran L, Zhang Q, Zhu J, Mi M (2015) Dihydromyricetin improves skeletal muscle insulin sensitivity by inducing autophagy via the AMPK-PGC-1α-Sirt3 signaling pathway. *Endocrine* **50**, 378-389.

[47] Liu J, Liu W, Ying H, Zhao W, Zhang H (2013) Analysis of microRNA expression profile induced by AICAR in mouse hepatocytes. *Gene* **512**, 364-372.

[48] Hölscher C (2012) Potential role of glucagon-like peptide-1 (GLP-1) in neuroprotection. *CNS Drugs* **26**, 871-882.

[49] Marwarha G, Raza S, Meiers C, Ghribi O (2014) Leptin attenuates BACE1 expression and amyloid-β genesis via the activation of SIRT1 signaling pathway. *Biochim Biophys Acta* **1842**, 1587-1595.

[50] Risner ME, Saunders AM, Altman JFB, Ormandy GC, Craft S, Foley IM, Zvartau-Hind ME, Hosford DA, Roses AD (2006) Efficacy of rosiglitazone in a genetically defined population with mild-to-moderate Alzheimer's disease. *Pharmacogenomics J* **6**, 246-254.

[51] Kim B, Sullivan KA, Backus C, Feldman EL (2011) Cortical neurons develop insulin resistance and blunted Akt signaling: A potential mechanism contributing to enhanced ischemic injury in diabetes. *Antioxid Redox Signal* **14**, 1829-1839.

[52] Chen Y, Zhou K, Wang R, Liu Y, Kwak Y-D, Ma T, Thompson RC, Zhao Y, Smith L, Gasparini L, Luo Z, Xu H, Liao FF (2009) Antidiabetic drug metformin (GlucophageR) increases biogenesis of Alzheimer's amyloid peptides via up-regulating BACE1 transcription. *Proc Natl Acad Sci U S A* **106**, 3907-3912.

[53] Kim D-H, Huh J-W, Jang M, Suh J-H, Kim T-W, Park J-S, Yoon S-Y (2012) Sitagliptin increases tau phosphorylation in the hippocampus of rats with type 2 diabetes and in primary neuron cultures. *Neurobiol Dis* **46**, 52-58.

[54] Malhotra A, Younesi E, Gurulingappa H, Hofmann-Apitius M, Rzhetsky A (2013) 'HypothesisFinder:' A strategy for the detection of speculative statements in scientific text. *PLoS Comput Biol* **9**, e1003117.

[55] Kirpichnikov D, McFarlane SI, Sowers JR (2002) Metformin: An update. *Ann Intern Med* **137**, 25-33.

[56] Talbot K, Wang H-Y, Kazi H, Han L-Y, Bakshi KP, Stucky A, Fuino RL, Kawaguchi KR, Samoyedny AJ, Wilson RS, Arvanitakis Z, Schneider JA, Wolf BA, Bennett DA, Trojanowski JQ, Arnold SE (2012) Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline. *J Clin Invest* **122**, 1316-1338.

[57] Zhao W-Q, De Felice FG, Fernandez S, Chen H, Lambert MP, Quon MJ, Krafft GA, Klein WL (2008) Amyloid beta oligomers induce impairment of neuronal insulin receptors. *FASEB J* **22**, 246-260.

[58] Kickstein E, Krauss S, Thornhill P, Rutschow D, Zeller R, Sharkey J, Williamson R, Fuchs M, Köhler A, Glossmann H, Schneider R, Sutherland C, Schweiger S (2010) Biguanide metformin acts on tau phosphorylation via mTOR/protein phosphatase 2A (PP2A) signaling. *Proc Natl Acad Sci U S A* **107**, 21830-21835.

[59] Patrone C, Eriksson O, Lindholm D (2014) Diabetes drugs and neurological disorders: New views and therapeutic possibilities. *Lancet Diabetes Endocrinol* **2**, 256-262.

[60] Li J, Deng J, Sheng W, Zuo Z (2012) Metformin attenuates Alzheimer's disease-like neuropathology in obese, leptin-resistant mice. *Pharmacol Biochem Behav* **101**, 564-574.

[61] Ouslimani N, Mahrouf M, Peynet J, Bonnefont-Rousselot D, Cosson C, Legrand A, Beaudeux JL (2007) Metformin reduces endothelial cell expression of both the receptor for advanced glycation end products and lectin-like oxidized receptor 1. *Metabolism* **56**, 308-313.

[62] Fu A, Eberhard CE, Screaton RA (2013) Role of AMPK in pancreatic beta cell function. *Mol Cell Endocrinol* **366**, 127-134.

[63] Picone P, Nuzzo D, Caruana L, Messina E, Barera A, Vasto S, Di Carlo M (2015) Metformin increases APP expression and processing via oxidative stress, mitochondrial dysfunction and NF-κB activation: Use of insulin to attenuate metformin's effect. *Biochim Biophys Acta* **1853**, 1046-1059.

[64] Barini E, Antico O, Zhao Y, Asta F, Tucci V, Catelani T, Marotta R, Xu H, Gasparini L (2016) Metformin promotes tau aggregation and exacerbates abnormal behavior in a mouse model of tauopathy. *Mol Neurodegener* **11**, 1.

[65] Buvat DR (2004) Use of metformin is a cause of vitamin B12 deficiency. *Am Fam Physician* **69**, 264; author repl 264, 266.

[66] Zhou Y, James I, Besner GE (2012) Heparin-binding epidermal growth factor–like growth factor promotes murine enteric nervous system development and enteric neural crest cell migration. *J Pediatr Surg* **47**, 1865-1873.

Summary

The basis of this work is systematic harvesting of relevant biological information from previously published research, encoded in BEL, to create disease-specific knowledge assemblies. By comparing and analyzing interactions and events of AD and T2DM, our work provides a broad and detailed mechanistic overview of comorbidity between the two diseases. For instance, we have been able to depict chains of causal molecular interactions that are involved in perturbing insulin signaling, which eventually give rise to AD phenotypes, in particular, formation of amyloid beta peptides and NFTs. Our work establishes a comorbid link between AD and T2DM at a more granular level because we have demonstrated functional roles of molecular entities which are often missing in previously published studies. Furthermore, our analysis has identified cross-talk among several signaling pathways such as insulin signaling, neurotrophin signaling, PI3K/AKT signaling, MTOR signaling and MAPK signaling, all of which are involved in the pathogenesis of AD and T2DM. Moreover, impairment in signaling pathways induced by altered activity of genes/proteins was rationalized by over- or under-expressed genes/proteins observed in omics data.

In addition to this, we have explored putative beneficial and harmful effects of Metformin, an FDA approved T2DM drug, in AD. Our analysis illustrates that, despite Metformin's capacity to alleviate AD-related dysfunctional activities, such as impaired synaptic transmission and neuronal damage, it can exacerbate the accumulation of amyloid beta peptides by increasing expression of BACE1 and PSEN1 and can promote neuroinflammation and oxidative stress. These findings call into question the potential for repurposing Metformin for therapeutic use in AD, described in previous studies. Furthermore, we hypothesize that Metformin could contribute to the comorbidity between AD and T2DM by illustrating AD-related events induced by Metformin.

Chapter 3

Data-driven modeling of knowledge assemblies in understanding comorbidity between Type 2 Diabetes Mellitus and Alzheimer's Disease

Introduction

In chapter 2, we presented a work that established comorbidity between AD and T2DM using published literature. However, the issue of literature bias, a consequence of recurrently investigated biological entities of known functions, is also inherited in the knowledge assemblies. For example, because APP and APOE are the focus of so much research in AD, a knowledge assembly for AD is likely to contain an over-representation of interactions involving these genes. This suggests that although literature-based modeling approaches enable the assembly of previously reported findings, insights thus generated have a certain degree of publication bias. In this chapter, using genomic data and omics data as the driving forces of our study, we conducted a publication-bias-free analysis in understanding comorbidity between AD and T2DM. Firstly, genomic data for were collected from curated databases such as GWAS catalog [145], GWAS central [146], dbSNP [147] and DisGeNET [148]. These were subjected to linkage disequilibrium (LD) analysis to identify additional genetic variants that have non-random association with variants initially collected. Secondly, after identifying shared variants between AD and T2DM, we prioritized the variants using PolyPhen-2 [149] and RegulomeDB [150]. Using the corresponding genes of the variants, we built a knowledge assembly representing their functional roles at molecular levels. Thirdly, analyzing omics data from Gene Expression Omnibus (GEO), we performed a differential gene expression meta-analysis to identify coherently perturbed genes in AD and T2DM. Lastly, we mapped these genes to our knowledge assembly to unravel their pleiotropic roles as a comorbid link between AD and T2DM.

Journal of Alzheimer's Disease xx (2020) x–xx
DOI 10.3233/JAD-200752
IOS Press

Data-Driven Modeling of Knowledge Assemblies in Understanding Comorbidity Between Type 2 Diabetes Mellitus and Alzheimer's Disease

Reagon Karki[a,b,*], Sumit Madan[a], Yojana Gadiya[a,b], Daniel Domingo-Fernández[a], Alpha Tom Kodamullil[a] and Martin Hofmann-Apitius[a,b]
[a]*Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany*
[b]*Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn-Aachen International Center for IT, Bonn, Germany*

Accepted 1 August 2020

Abstract.
Background: Recent studies have suggested comorbid association between Alzheimer's disease (AD) and type 2 diabetes mellitus (T2DM) through identification of shared molecular mechanisms. However, the inference is pre-dominantly literature-based and lacks interpretation of pre-disposed genomic variants and transcriptomic measurables.
Objective: In this study, we aim to identify shared genetic variants and dysregulated genes in AD and T2DM and explore their functional roles in the comorbidity between the diseases.
Methods: The genetic variants for AD and T2DM were retrieved from GWAS catalog, GWAS central, dbSNP, and DisGeNet and subjected to linkage disequilibrium analysis. Next, shared variants were prioritized using RegulomeDB and Polyphen-2. Afterwards, a knowledge assembly embedding prioritized variants and their corresponding genes was created by mining relevant literature using Biological Expression Language. Finally, coherently perturbed genes from gene expression meta-analysis were mapped to the knowledge assembly to pinpoint biological entities and processes and depict a mechanistic link between AD and T2DM.
Results: Our analysis identified four genes (i.e., *ABCG1*, *COMT*, *MMP9*, and *SOD2*) that could have dual roles in both AD and T2DM. Using cartoon representation, we have illustrated a set of causal events surrounding these genes which are associated to biological processes such as oxidative stress, insulin resistance, apoptosis and cognition.
Conclusion: Our approach of using data as the driving force for unraveling disease etiologies eliminates literature bias and enables identification of novel entities that serve as the bridge between comorbid conditions.

Keywords: Alzheimer's disease, comorbidity, systems biology, type 2 diabetes mellitus

Uncorrected Author Proof

INTRODUCTION

In recent years, comorbidities are inspected with a different perspective. The new route in understanding possible comorbidities has changed from classical approaches that use magnitude, severity, patterns, and burden to comparing disease associated events, pathways, and maps [1, 2]. By establishing comorbid

*Correspondence to: Reagon Karki, Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Room: C3-242, Schloss Birlinghoven, Konrad-Adenauer-Straße, 53754 Sankt Augustin, Germany. Tel.: +49 2241 14 2294: E-mail: reagon.karki@scai.fraunhofer.de.

ISSN 1387-2877/20/$35.00 © 2020 – IOS Press and the authors. All rights reserved
This article is published online with Open Access and distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC 4.0).

associations from assessment scores of these aspects, classical approaches fail to explain the biology underlying diseases. Hence, biological entities such as genes, proteins, and miRNAs and their involvement in biological processes and pathways have been studied to unravel insights about comorbidity.

The possible association between type 2 diabetes mellitus (T2DM) and Alzheimer's disease (AD) has enticed significant interest from the scientific community since the identification of typical events of T2DM in AD and vice-versa. For instance, the brains of AD patients are reported to exhibit T2DM-related mechanisms including impaired insulin signaling, insulin resistance, and impaired glucose metabolism [3]. Moreover, hyperphosphorylated microtubule associated protein tau (MAPT) leading to formation of neurofibrillary tangles (NFTs), one of the hallmarks of AD, is a consequence of abnormal glycogen synthase kinase 3 beta activity in the insulin signaling pathway [4]. On the other hand, the presence of abnormally processed islet amyloid polypeptide in pancreas of T2DM patients mimics amyloid-β protein precursor (AβPP)-derived deleterious amyloid-β (Aβ) in AD brains [5]. In addition to these, the comorbid link between T2DM and AD has been established through several studies reporting shared biological processes such as oxidative stress, mitochondrial dysfunction, inflammation, and advanced glycation end products [6].

While most of the speculations are based on individual experiments, studies, or review articles, the putative mechanisms explaining the comorbidity are still unknown. To address this issue, disease-specific knowledge assemblies are created by systematic retrieval of biological information from literature and compared for identifying shared pathophysiological mechanisms. In this regard, Kodamullil et al. (2015) have undertaken a systems biology approach to create cause-and-effect models and proposed single nucleotide polymorphism (SNP)-based mechanisms as the link between the diseases. This is one of the first and few studies that mechanistically depicts and compares disease etiologies [7]. A broader scenario representing mechanistic crosstalk between several pathways such as insulin signaling, neurotrophin signaling, inflammatory regulators, and MTOR signaling in AD and T2DM was demonstrated in our previous work [8]. Interestingly, we have also suggested that metformin, an FDA approved T2DM drug, could be one of the risk factors for developing AD in old age of the diabetic patients. Through this study, the consideration of metformin in drug repositioning in AD has been questioned by depicting the role of metformin in contributing to augment characteristic features of AD such as neuroinflammation, formation of Aβ, and NFTs. Therefore, the hypothesis of drug-induced comorbidity cannot be ruled out. In this context, prolonged use of anti-psychotic drugs has been previously reported to induce symptoms of Parkinson's disease (PD) [9–11]. The authors have rationalized this assumption by identifying blocked dopamine receptors and calcium channels by the drugs, both of which are impaired in PD. However, the postulation about this aspect of drugs in inducing a disease as a side-effect is still at its infancy.

The prevalence of study bias, which eventually leads to literature bias, is due to the fact that proteins with known biomedical functions and associated signaling pathways are studied recurrently [12, 13]. And because knowledge assemblies massively depend on literature resource, they inherit pre-existing bias. Therefore, chances are higher that literature aided inferences could represent biased knowledge. Taking this into consideration, Naz et al. (2017) have analyzed genomic data and performed functional assessment of prioritized SNPs using literature to depict stress-induced comorbid association in AD and PD [14]. This approach not only eliminates biasedness of over-representation of well-known biological entities and processes, but also identifies new genes and associated events which can serve as putative drug targets and drugable mechanisms. In this study, we have implemented a similar strategy in deciphering the comorbid link between T2DM and AD. The genomic data (i.e., SNPs) for AD and T2DM were fetched from curated public databases and subjected to linkage disequilibrium (LD) analysis. After filtering for shared SNPs in both diseases and prioritizing them based on their relevance to the diseases, we constructed cause-and-effect computable, network models using Biological Expression Language (BEL) [15]. The language enables conversion of unstructured textual information from literature into structured computer-readable triples (i.e., subject-predicate-object). The parsing and compilation of several triples after syntactic and semantic validation generates network models, which are also known as knowledge assemblies. Next, we added the dimension of high-throughput data as the driving force of our analysis by mapping differentially expressed genes to our knowledge assemblies. Finally, a mechanistic graph tailored by analysis of genomic and transcriptomic data was created from the knowledge assembly to explain the comorbid link between T2DM and AD.

Uncorrected Author Proof

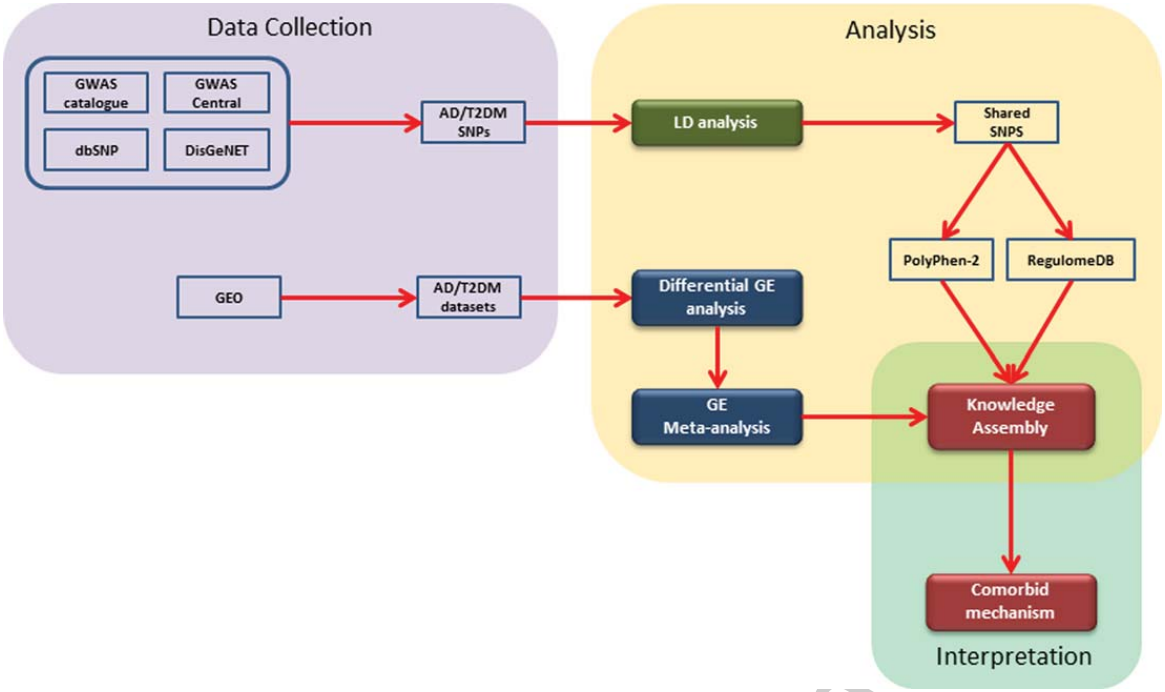

Fig. 1. A schematic representation of the implemented workflow: The steps involved are 1) collection of genomic and gene expression data from open and freely accessible databases; 2) analysis of data using available tools and packages; and 3) construction of literature derived knowledge assembly and comorbid interpretation.

MATERIALS AND METHODS

Firstly, a knowledge assembly embedding prioritized SNP was created from the literature. To this knowledge assembly, we mapped expression profiles from our gene expression analysis. Finally, we filtered the knowledge assembly with those genes which were consistently perturbed. The overall methodology implemented in this study can be divided into 1) data collection, 2) data analysis, and 3) interpretation. Firstly, we collected SNPs and gene expression (GE) datasets for AD and T2DM from freely accessible public databases (i.e., GWAS Catalog [16], GWAS Central [17], dbSNP [18], and DisGeNET [19]). Secondly, for SNP data, we conducted LD analysis followed by identification of shared SNPs and their prioritization using Polyphen-2 [20] and RegulomeDB [21]. Likewise, for GE datasets, we performed differential GE analysis followed by meta-analysis of AD and T2DM datasets. Lastly, we built a literature-derived knowledge assembly representing the results of the SNP analysis and mapped expression profiles of genes from the meta-analysis. A schematic diagram illustrating the methodology is shown in Fig. 1 and described in detail in the following sections.

Retrieval of AD and T2DM SNPs from curated SNP databases

We retrieved a total of 1,130, 1,516, 1,420, and 1,062 SNPs associated to AD from GWAS Catalog, GWAS Central, dbSNP, and DisGeNET, respectively. Similarly, we extracted 1,791, 1,069, 1,865, and 1,522 SNPs associated to T2DM from GWAS Catalog, GWAS Central, dbSNP, and DisGeNET, respectively. To ensure the accuracy of our search, we queried dbSNP, GWAS central, and DisGeNET with corresponding Medical Subject Headings (MeSH) identifiers of AD (i.e., D000544) and T2DM (i.e., D003924). Likewise, we used Experimental Ontology Factor (EFO) identifiers of AD (i.e., EFO_0000249) and T2DM (i.e., EFO_0001360) for GWAS Catalog.

Linkage disequilibrium analysis and SNP prioritization

Using a total of 5,128 and 6,247 SNPs associated to AD and T2DM, respectively, we performed a LD analysis using the R-package *haploR* [22]. The function *queryHaploreg* was used with the default r^2 threshold of 0.8 to perform the analysis. This yielded 77,486 SNPs in AD and 130,807 SNPs in T2DM. Out

Uncorrected Auth

of these, 3,572 SNPs were shared between the diseases. Next, depending on whether the SNPs occur in coding or non-coding region of the gene, we used two databases to functionally annotate these shared SNPs. The impact of the SNPs located in the coding region and the resulting amino acid mutation along with the prediction, either benign or possibly damaging, was assessed using Polyphen-2. Likewise, the assessment of the SNPs in the regulatory region (non-coding) was performed using the function *queryRegulome* from R-package *haploR*. Subsequently, we prioritized them using RegulomeDB scores based on current ENCODE releases, Chromatin States from the Roadmap Epigenome Consortium as well as updates to DNase footprinting, Position Weight Matrix for TF binding, and DNA Methylation, and ENSEMBL SNP's functional consequences [21, 23].

Literature corpus and cause-and-effect model using Biological Expression Language

The functional annotation of SNPs using Polyphen-2 and RegulomeDB helps in prioritization of SNPs. Nonetheless, these databases lack their putative roles in a disease context. In this study, we aimed at depicting mechanistic causal graphs embedding prioritized SNPs and their corresponding genes. This was achieved by building a comprehensive knowledge assembly using MEDLINE as the source of literature. The MeSH terms "*Alzheimer Disease*" and "*Diabetes Mellitus, Type 2*" were used to query PubMed (Date:02-12-2019) to create separate literature corpus of both diseases. The total number of articles for AD and T2DM were 90,215 and 127,020, respectively. Furthermore, through text mining, we created literature corpora that only contained shared SNPs and genes from LD analysis. The new corpus corresponding to AD and T2DM had a total of 14,293 and 9,032 articles, respectively. Next, we used BEL to capture causal and correlative relationships between the entities from the corpora. The language serves as an efficient platform to create computable knowledge assemblies by compiling relationships which are formulated in the form of triples. The conversion of regular text to BEL was assisted by BELIEF, a semi-automatic workflow to systematically extract BEL relationships from the corpus [24]. The outputs of the BELIEF workflow were manually curated to ensure high quality of the BEL relationships and then compiled using PyBEL for visualization [25].

Meta-analysis of gene expression datasets

In this study, our objective is to perform functional assessment of shared SNPs between AD and T2DM with the help of literature derived knowledge assemblies. In the Introduction section, we have already mentioned the possible bias that results from a purely literature-based construction of knowledge assemblies. Therefore, in order to tackle this issue, we mapped and investigated genes with consistent patterns of perturbed expressions to the knowledge assembly as such genes are more likely to be important in disease pathophysiology. A total of 14 GE datasets, 7 each for AD and T2DM, were selected from GEO (Gene Expression Omnibus). The selection of the datasets was done based on the criterion that the samples must be from humans (i.e., patients) diagnosed with AD or T2DM. Moreover, we did not consider datasets that used cell lines, induced medical conditions, animal models and modified genes or environments for expression analysis. The datasets were analyzed with GEO2R tool to identify differentially expressed genes in both diseases [26]. However, because expression patterns of the same disease are inconsistent and non-reproducible [27, 28], we performed a meta-analysis of the AD and T2DM GE datasets independently. This was achieved by using *MetaVolcanoR*, an R package with an algorithm based on voting approach and *p*-values of differentially expressed genes [29]. This allowed us to identify consistent patterns of perturbed gene expression across all the datasets. A brief description of each of the datasets is provided in Supplementary File 1.

RESULTS

Linkage disequilibrium analysis

The distribution analysis of 3,572 shared SNPs revealed that chromosome 1 had the highest number of SNPs, i.e., 495, followed by chromosome 17 (295 SNPs) and chromosome 8 (289 SNPs). The distribution of SNPs over all the chromosomes is shown in Supplementary File 2. The shared SNPs were mapped to 236 genes and the top 5 genes with the highest number of SNPS were lipoprotein lipase (*LPL*) (CHR 8, 153 SNPs), ubiquitin conjugating enzyme E2 D3 (*UBE2D3*) (CHR 4, 128 SNPs), leptin receptor (*LEPR*) (CHR 1, 116 SNPs), FTO alpha-ketoglutarate dependent dioxygenase (*FTO*) (CHR 16, 94 SNPs), and EF-hand calcium binding domain

Uncorrected Author Proof

5 (*EFCAB5*) (CHR 17, 86 SNPs). The full list of number of SNPs per each gene is provided in Supplementary File 3.

Assessment of SNPs with Polyphen-2 and RegulomeDB

A total of 64 SNPs, mapped to 50 genes, were identified by Polyphen-2 to be responsible for amino acid substitutions in their corresponding proteins. Out of these, 50 mutations were predicted to be benign while the remaining 14 mutations were predicted to be possibly damaging. Interestingly, mutations in the few well-characterized genes in AD and T2DM such as apolipoprotein E (*APOE*), brain derived neurotrophic factor (*BDNF*), and insulin receptor substrate 1 (*IRS1*) were classified as "possibly damaging". The full list of Polyphen-2 output is provided in Supplementary File 4. Likewise, a total of 127 SNPs, mapped to 52 genes, were identified by RegulomeDB to be located in the functional region of their corresponding genes. This was indicated by the scores ranging between 1a and 1f. Genes such as *APOE*, translocase of outer mitochondrial membrane 40 (*TOMM40*), and interleukin 6 (*IL6*) were among the examples for the genes that were mapped to the 127 SNPs. The full list of RegulomeDB output is provided in Supplementary File 5.

Results from meta-analysis of GE datasets

The meta-analysis of AD datasets showed 206 genes to exhibit consistent patterns of perturbed expression, where 49 genes were underexpressed and 157 genes were overexpressed. Similarly, in T2DM, a total of 142 genes regulated persistently, with 13 genes showing downregulation and 129 genes that were consistently upregulated. Out of these, 3 genes, i.e., interferon gamma inducible protein 16 (*IFI16*), syntrophin beta 2 (*SNTB2*), and laminin subunit alpha 4 (*LAMA4*) were found to be overexpressed in meta-analyses of AD and T2DM. The full list of differentially expressed genes and plots showing expression patterns of each datasets are provided in Supplementary Files 6 and 7, respectively. The implementation of GE meta-analysis after differential GE analysis is justified by our findings that the number of coherently perturbed genes reduced with increasing number of GE datasets. This implies the ability of GE meta-analysis to yield robustness and convergence of expression patterns.

Comorbidity in AD and T2DM explained by mechanistic BEL graphs

The knowledge assemblies representing AD and T2DM were combined to investigate the role of shared SNPs and their corresponding genes along with consistently perturbed genes from our meta-analyses. The merged network had a total of 692 nodes and 1,793 edges. The top 5 biological processes based on highest degree of node centrality were insulin resistance, inflammatory response, aggregation of Aβ, apoptotic process, and oxidative stress. Similarly, cystatin C (*CST3*), *BDNF*, peroxisome proliferator activated receptor gamma (*PPARG*), *MAPT*, and *LEPR* were the top 5 genes in the network. We mapped persistently perturbed genes from the meta-analyses to the network and used them as driving force of our comorbid analysis. The rationale supporting this implementation are 1) abnormal expression of genes and their activities mutilate biological processes and pathways and thus are responsible for manifesting disease characteristics, and 2) it overcomes the risk of representing biased knowledge. A mechanistic graph embedding corresponding genes of our SNP analysis and abnormally expressed genes is shown in Fig. 2 It had a total of 41 nodes and 45 edges and comprised of 4 genes (i.e., cholinergic receptor nicotinic alpha 3 subunit (*CHRNA3*) (CHR 15, 6 SNPs), catechol-O-methyltransferase (*COMT*) (CHR 22, 4 SNPs), nuclear receptor subfamily 1 group H member 3 (*NR1H3*) (CHR 11, 13 SNPs), and transforming growth factor beta 1 (*TGFB1*) (CHR 19, 4 SNPs) sharing 27 SNPs between AD and T2DM.

As shown in Fig. 2, *COMT* is known to influence synaptic plasticity and dopamine metabolism, both of which are associated with cognition. In this context, two point mutations (i.e., Val108Met and rs4680 –> Val158Met) in this gene were identified to be predictors of cognition scores in AD patients through independent studies [30, 31]. Interestingly, the latter mutation along with rs4646312 in *COMT* has been associated with T2DM [32]. Also, the 900delC variant form in *COMT* correlates to chronic renal insufficiency in T2DM [33]. Along the same lines, C47T variant in superoxide dismutase 2 (*SOD2*) is associated with cognition [34]. Likewise, ATP binding cassette subfamily G member 1 (*ABCG1*), which is upregulated by *NR1H3* [35], has been linked with T2DM because of its involvement in obesity and lipid metabolism [36, 37]. In AD, *ABCG1* is reported to inhibit the process of formation of Aβ through inhibition of

Uncorrected Author Proof

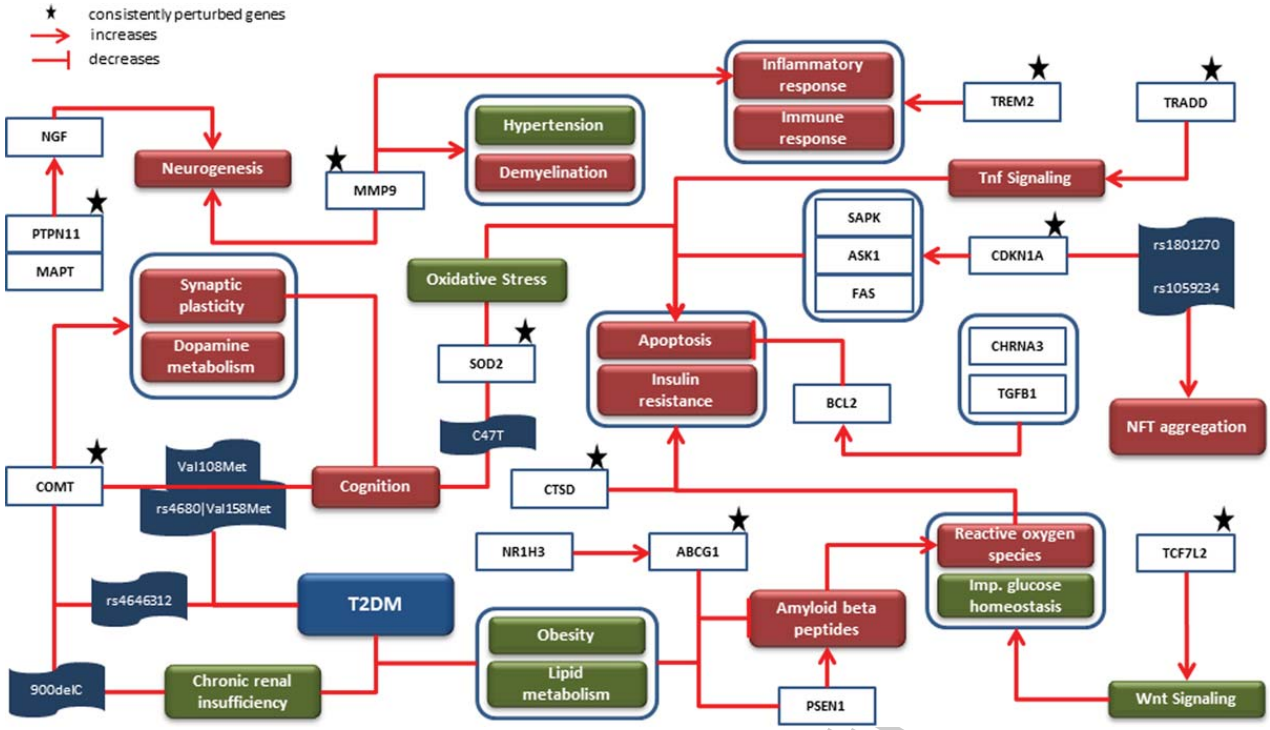

Fig. 2. Mechanistic comorbid association: The cartoon diagram represents causal interactions between entities and events led by prioritized SNPs, their corresponding genes and consistently dysregulated genes. Genes such as *COMT*, *MMP9*, *ABCG1*, and *SOD2* were identified to contribute to biological processes such as oxidative stress, apoptosis and cognition, all of which are associated with T2DM and AD. Since the onset of T2DM takes place earlier than AD, this diagram is inclined toward AD-related phenomena (i.e., NFT aggregation, synaptic plasticity, etc.) which are influenced by downstream T2DM-related genes.

presenilin 1 (PSEN1) [38]. The amyloid hypothesis in AD considers PSEN1 as one of the two important enzymes that is responsible for abnormal cleaving of AβPP [39], the other being beta-secretase 1 (BACE1) [40]. The accumulated Aβ leading to production of reactive oxygen species (ROS) which further increases oxidative stress to consequently trigger apoptosis is well understood through several studies in AD [41, 42]. Similarly, transcription factor 7 like 2 (*TCF7L2*)-activated Wnt signaling has been reported to generate ROS and impair glucose homeostasis [43]. The excess of ROS thus produced is detrimental as it results in insulin resistance [44]. Moreover, another study has identified dysfunctional cathepsin D (*CTSD*) to increase both insulin resistance and apoptosis [45, 46]. In addition, *SOD2* induced oxidative stress [47], *TNFRSF1A* associated via death domain (*TRADD*) activated Tnf signaling, and cyclin dependent kinase inhibitor 1A (*CDKN1A*) activated mitogen-activated protein kinase 9 (*MAPK9*) [48], mitogen-activated protein kinase kinase kinase 5 (*MAP3K5*), and Fas cell surface death receptor (*FAS*) also lead to apoptosis [49]. In contrast, the suppression of apoptosis takes place through *CHRNA3* [50] and *TGFB1* activated BCL2 apoptosis regulator (*BCL2*) [51]. The identification of two SNPs (i.e., rs1801270 and rs1059234) in *CDKN1A* positively correlated with NFT aggregation in AD patients [49]. The binding of protein tyrosine phosphatase non-receptor type 11 (*PTPN11*) and *MAPT* promotes neurogenesis by activating nerve growth factor [52]. Besides, matrix metallopeptidase 9 (*MMP9*) is also suggested to play a part in neurogenesis and other biological processes such as hypertension, demyelination, and inflammatory response [53, 54]. Lastly, triggering receptor expressed on myeloid cells 2 (*TREM2*) activity is found to influence inflammatory and immune response [55].

DISCUSSION

In this study, we formulated an integrative approach of combining data and knowledge to unravel new insights about the possible association between AD and T2DM. Our data-driven modeling of knowledge assemblies represents highly

Uncorrected Art

specialized knowledge on comorbidities. While most of the data analytics workflow end up in gene set enrichment analysis, our approach has opened up a new avenue of mechanism-centric interpretation of data. We have used two different data modalities to guide the extraction process of relevant literature knowledge. Firstly, we identified shared SNPs between AD and T2DM from curated resources and built a knowledge assembly around prioritized SNPs and their corresponding genes. Although literature can also be used as a source of SNP information, it is important to note that we have considered only curated databases for retrieval of SNPs. This decision can be explained by the fact that curated databases ensure association between a SNP and a disease with a given statistical significance (i.e., *p*-value). Unlike curated databases, some SNPs mentioned in the literature might not have any association with a disease because the statistical power of association is below par [56–59]. Therefore, by including such SNPs, we would be adding possible false positives in our analysis and, thus, diminishing the quality of the results. Secondly, as we are aware of the literature bias in knowledge assemblies, we identified consistently perturbed genes by conducting a GE meta-analysis and used these signals to mechanistically link AD and T2DM. Our results illustrate that genes such as *COMT*, *MMP9*, *SOD2*, and *ABCG1*, which do not belong to the realm of well-known genes in AD and T2DM, are involved in important biological processes of both diseases. This suggests dysfunctional activities of these genes could be the bridge between these diseases. Moreover, our findings endorse and strengthen the proposition of AD and T2DM comorbidity suggested by epidemiological, preclinical, and pathophysiology studies by identifying novel genes.

The genetic variants of AD and T2DM amassed in our study are readily explorable and bear the potential to yield new insights. For instance, genomic loci dependent SNPs can be functionally assessed to uncover their roles in the underlying comorbid mechanisms. This would enable identification of "genomic hotspots" that are closely associated to AD and T2DM. However, this study does not address this aspect due to time constraints and it is out-of-scope of our objectives. Also, we have not considered the role of epigenetic modifications in the comorbid association between AD and T2DM. Nonetheless, our knowledge assemblies can be used as the starting point for assimilating epigenetic modifications concerning AD and T2DM.

ACKNOWLEDGMENTS

This work was funded by Bundesministerium für Bildung und Forschung (BMBF, e:Med initiative COMMITMENT (grant number: 01ZX1904C)). The authors would also like to acknowledge the financial support from the B-IT foundation that sponsors part of the academic work in our department.

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/20-0752r1).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-200752.

REFERENCES

[1] Haregu TN, Oldenburg BF, Setswe G, Elliott J (2012) Perspectives, constructs and methods in the measurement of multimorbidity and comorbidity: A critical review. *Internet J Epidemiol* **10**, 1-11.

[2] Pugazhenthi S, Qin L, Reddy PH (2017) Common neurodegenerative pathways in obesity, diabetes, and Alzheimer's disease. *Biochim Biophys Acta Mol Basis Dis* **1863**, 1037-1045.

[3] de la Monte SM, Wands JR (2008) Alzheimer's disease is type 3 diabetes—evidence reviewed. *J Diabetes Sci Technol* **2**, 1101-1113.

[4] de la Monte SM, Chen GJ, Rivera E, Wands JR (2003) Neuronal thread protein regulation and interaction with microtubule-associated proteins in SH-Sy5y neuronal cells. *Cell Mol Life Sci C* **60**, 2679-2691.

[5] Yashin AI, Fang F, Kovtun M, Wu D, Duan M, Arbeev K, Akushevich I, Kulminski A, Culminskaya I, Zhbannikov I, Tashkin A, Stallard E, Ukraintseva S (2018) Hidden heterogeneity in Alzheimer's disease: Insights from genetic association studies and other analyses. *Exp Gerontol* **107**, 148-160.

[6] Akter K, Lanza EA, Martin SA, Myronyuk N, Rua M, Raffa RB (2011) Diabetes mellitus and Alzheimer's disease: Shared pathology and treatment? *Br J Clin Pharmacol* **71**, 365-376.

[7] Kodamullil AT, Younesi E, Naz M, Bagewadi S, Hofmann-Apitius M (2015) Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis. *Alzheimers Dement* **11**, 1329-1339.

[8] Karki R, Kodamullil AT, Hofmann-Apitius M (2017) Comorbidity analysis between Alzheimer's disease and type 2 diabetes mellitus (T2DM) based on shared pathways and the role of T2DM drugs. *J Alzheimers Dis* **60**, 721-731.

[9] Bohlega SA, Al-Foghom NB (2013) Drug-induced Parkinson's disease. A clinical review. *Neurosci* **18**, 215-221.

[10] Shin HW, Chung SJ (2012) Drug-induced Parkinsonism. *J Clin Neurol* **8**, 15-21.

[11] Bashford G, Bradd P (1996) Drug-induced Parkinsonism associated with dysphagia and aspiration: A brief report. *J Geriatr Psychiatry Neurol* **9**, 133-135.

Uncorrected Author Proof

[12] Schaefer MH, Serrano L, Andrade-Navarro MA (2015) Correcting for the study bias associated with protein–protein interaction measurements reveals differences between protein degree distributions from different cancer types. *Front Genet* **6**, 260.

[13] Charitou T, Bryan K, Lynn DJ (2016) Using biological networks to integrate, visualize and analyze genomics data. *Genet Sel Evol* **48**, 27.

[14] Naz M, Younesi E, Hofmann-Apitius M (2017) Systematic analysis of GWAS data reveals genomic hotspots for shared mechanisms between neurodegenerative diseases. *J Alzheimers Dis Park* **7**, 460-2161.

[15] Slater T, Song D (2012) Saved by the BEL: Ringing in a common language for the life sciences. *Drug Discov World Fall* **2012**, 75-80.

[16] Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F, Parkinson H (2019) The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012.

[17] Beck T, Shorter T, Brookes AJ (2020) GWAS central: A comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies. *Nucleic Acids Res* **48**, D933-D940.

[18] Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000) dbSNP: A database of single nucleotide polymorphisms. *Nucleic Acids Res* **28**, 352-355.

[19] Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI (2019) The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* **48**, D845-D855.

[20] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249.

[21] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-1797.

[22] Zhbannikov IY, Arbeev K, Ukraintseva S, Yashin AI (2017) *haploR*: An R package for querying web-based annotation tools. *F1000Res* **6**, 97.

[23] Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K, García Girón C, Hourlier T, Howe K, Kähäri A, Kokocinski F, Martin FJ, Murphy DN, Nag R, Ruffier M, Schuster M, Tang YA, Vogel JH, White S, Zadissa A, Flicek P, Searle SMJ (2016) The Ensembl gene annotation system. *Database (Oxford)* **2016**, baw093.

[24] Fluck J, Madan S, Ansari S, Szostak J, Hoeng J, Zimmermann M, Hofmann-Apitius M, Peitsch MC (2014) BELIEF-A semiautomatic workflow for BEL network creation. In *Proceedings of the 6th International Symposium on Semantic Mining in Biomedicine*, pp. 109-113.

[25] Hoyt CT, Konotopez A, Ebeling C (2018) PyBEL: A computational framework for biological expression language. *Bioinformatics* **34**, 703-704.

[26] Clough E, Barrett T (2016) The gene expression omnibus database. *Methods Mol Biol* **1418**, 93-110.

[27] Jafari P, Azuaje F (2006) An assessment of recently published gene expression data analyses: Reporting experimental design and statistical factors. *BMC Med Inform Decis Mak* **6**, 27.

[28] Ntzani EE, Ioannidis JP (2003) Predictive ability of DNA microarrays for cancer outcomes and correlates: An empirical assessment. *Lancet* **362**, 1439-1444.

[29] Prada C, Lima D, Nakaya H (2020) MetaVolcanoR: Gene expression meta-analysis visualization tool. *R Packag. version 1.1*.

[30] Lanni C, Garbin G, Lisa A, Biundo F, Ranzenigo A, Sinforiani E, Cuzzoni G, Govoni S, Ranzani GN, Racchi M (2012) Influence of COMT Val158Met polymorphism on Alzheimer's disease and mild cognitive impairment in Italian patients. *J Alzheimers Dis* **32**, 919-926.

[31] Perkovic MN, Strac DS, Tudor L, Konjevod M, Erjavec GN, Pivac N (2018) Catechol-O-methyltransferase, cognition and Alzheimer's disease. *Curr Alzheimer Res* **15**, 408-419.

[32] Xiu L, Lin M, Liu W, Kong D, Liu Z, Zhang Y, Ouyang P, Liang Y, Zhong S, Chen C, Jin X, Fan X, Qin J, Zhao X, Rao S, Ding Y (2015) Association of DRD3, COMT, and SLC6A4 gene Polymorphisms with type 2 diabetes in Southern Chinese: A hospital-based case-control study. *Diabetes Technol Ther* **17**, 580-586.

[33] Prasad P, Kumar KM, Ammini AC, Gupta A, Gupta R, Thelma BK (2008) Association of dopaminergic pathway gene polymorphisms with chronic renal insufficiency among Asian Indians with type-2 diabetes. *BMC Genet* **9**, 26.

[34] Gamarra D, Elcoroaristizabal X, Fernández-Martínez M, de Pancorbo MM (2015) Association of the C47T polymorphism in SOD2 with amnestic mild cognitive impairment and Alzheimer's disease in carriers of the APOEε_4 allele. *Dis Markers* **2015**, 746329.

[35] Natunen T, Martiskainen H, Sarajärvi T, Helisalmi S, Pursiheimo JP, Viswanathan J, Laitinen M, Mäkinen P, Kauppinen T, Rauramaa T, Leinonen V, Alafuzoff I, Haapasalo A, Soininen H, Hiltunen M (2013) Effects of NR1H3 genetic variation on the expression of liver X receptor α and the progression of Alzheimer's disease. *PLoS One* **8**, e80700.

[36] Choromanska B, Mysliwiec P, Hady HR, Dadan J, Mysliwiec H, Bonda T, Chabowski A, Miklosz A (2019) The implication of adipocyte ATP-binding cassette A1 and G1 transporters in metabolic complications of obesity. *J Physiol Pharmacol* **70**, doi: 10.26402/jpp.2019.1.14

[37] Churm R, Caplin S, Barry J, Davies JS, Stephens JW, Prior SL (2019) Acyl-ghrelin mediated lipid retention and inflammation in obesity-related type 2 diabetes. *Mol Cell Endocrinol* **481**, 8-13.

[38] Sano O, Tsujita M, Shimizu Y, Kato R, Kobayashi A, Kioka N, Remaley AT, Michikawa M, Ueda K, Matsuo M (2016) ABCG1 and ABCG4 suppress γ-secretase activity and amyloid β production. *PLoS One* **11**, e0155400.

[39] Selkoe DJ (1994) Cell biology of the amyloid beta-protein precursor and the mechanism of Alzheimer's disease. *Annu Rev Cell Biol* **10**, 373-403.

[40] Cai H, Wang Y, McCarthy D, Wen H, Borchelt DR, Price DL, Wong PC (2001) BACE1 is the major beta-secretase for generation of Abeta peptides by neurons. *Nat Neurosci* **4**, 233-234.

[41] Cai Z, Zhao B, Ratka A (2011) Oxidative stress and β-amyloid protein in Alzheimer's disease. *Neuromolecular Med* **13**, 223-250.

[42] Loh KP, Huang SH, De Silva R, Tan BK, Zhu YZ (2006) Oxidative stress: Apoptosis in neuronal injury. *Curr Alzheimer Res* **3**, 327-337.

Uncorrected Author Proof

[43] Yoon JC, Ng A, Kim BH, Bianco A, Xavier RJ, Elledge SJ (2010) Wnt signaling regulates mitochondrial physiology and insulin sensitivity. *Genes Dev* **24**, 1507-1518.

[44] Zhai L, Ballinger SW, Messina JL (2011) Role of reactive oxygen species in injury-induced insulin resistance. *Mol Endocrinol* **25**, 492-502.

[45] Nowak C, Sundström J, Gustafsson S, Giedraitis V, Lind L, Ingelsson E, Fall T (2016) Protein biomarkers for insulin resistance and type 2 diabetes risk in two large community cohorts. *Diabetes* **65**, 276-284.

[46] Di YQ, Han XL, Kang XL, Wang D, Chen CH, Wang JX, Zhao XF (2020) Autophagy triggers CTSD (cathepsin D) maturation and localization inside cells to promote apoptosis. *Autophagy*, doi: 10.1080/15548627.2020.1752497

[47] Pias EK, Ekshyyan OY, Rhoads CA, Fuseler J, Harrison L, Aw TY (2003) Differential effects of superoxide dismutase isoform expression on hydroperoxide-induced apoptosis in PC-12 cells. *J Biol Chem* **278**, 13294-13301.

[48] Zhao M, Cribbs DH, Anderson AJ, Cummings BJ, Su JH, Wasserman AJ, Cotman CW (2003) The induction of the TNFalpha death domain signaling pathway in Alzheimer's disease brain. *Neurochem Res* **28**, 307-318.

[49] Yates SC, Zafar A, Rabai EM, Foxall JB, Nagy S, Morrison KE, Clarke C, Esiri MM, Christie S, Smith AD, Nagy Z (2015) The effects of two polymorphisms on p21cip1 function and their association with Alzheimer's disease in a population of European descent. *PLoS One* **10**, e0114050.

[50] Zhang XL, Qi XL, Ren JM, Wu CX, Guan ZZ (2013) Effects of α3 neuronal nicotinic acetylcholine receptor on cell apoptosis and p38 MAPK signal transduction pathway in SH-SY5Y cells. *Zhonghua Bing Li Xue Za Zhi* **42**, 116-120.

[51] Tan J, Town T, Placzek A, Kundtz A, Yu H, Mullan M (1999) Bcl-X(L) inhibits apoptosis and necrosis produced by Alzheimer's beta-amyloid1-40 peptide in PC12 cells. *Neurosci Lett* **272**, 5-8.

[52] Kim Y, Liu G, Leugers CJ, Mueller JD, Francis MB, Hefti MM, Schneider JA, Lee G (2019) Tau interacts with SHP2 in neuronal systems and in Alzheimer's disease brains. *J Cell Sci* **132**, jcs229054.

[53] Kaminari A, Tsilibary EC, Tzinia A (2018) A new perspective in utilizing MMP-9 as a therapeutic target for Alzheimer's disease and type 2 diabetes mellitus. *J Alzheimers Dis* **64**, 1-16.

[54] Kostov K, Blazhev A, Atanasova M, Dimitrova A (2016) Serum concentrations of endothelin-1 and matrix metalloproteinases-2, -9 in pre-hypertensive and hypertensive patients with type 2 diabetes. *Int J Mol Sci* **17**, 1182.

[55] Misra A, Chakrabarti SS, Gambhir IS (2018) New genetic players in late-onset Alzheimer's disease: Findings of genome-wide association studies. *Indian J Med Res* **148**, 135-144.

[56] Sajovic J, Cilenšek I, Mankoč S, Tajnšek Š, Kunej T, Petrovič D, Petrovič MG (2019) Vascular endothelial growth factor (VEGF)-related polymorphisms rs10738760 and rs6921438 are not risk factors for proliferative diabetic retinopathy (PDR) in patients with type 2 diabetes mellitus (T2DM). *Bosn J Basic Med Sci* **19**, 94.

[57] Završnik M, Letonja J, Makuc J, Šeruga M, Cilenšek I, Petrovič D (2018) Interleukin-4 (IL4) -590C/T (rs2243250) gene polymorphism is not associated with diabetic nephropathy (DN) in Caucasians with type 2 diabetes mellitus (T2DM). *Bosn J Basic Med Sci* **18**, 347-351.

[58] Bey K, Wolfsgruber S, Karaca I, Wagner H, Lardenoije R, Becker J, Milz E, Kornhuber J, Peters O, Frölich L, Hüll M, Rüther E, Wiltfang J, Riedel-Heller S, Scherer M, Jessen F, Maier W, van den Hove DL, Rutten BP, Wagner M, Ramirez A (2016) No association of the variant rs11887120 in DNMT3A with cognitive decline in individuals with mild cognitive impairment. *Epigenomics* **8**, 593-598.

[59] Zeng F, Deng YP, Yi X, Cao HY, Zou HQ, Wang X, Liang CR, Wang YR, Zhang LL, Gao CY, Xu ZQ, Lian Y, Wang L, Zhou XF, Zhou HD, Wang YJ (2013) No association of SORT1 gene polymorphism with sporadic Alzheimer's disease in the Chinese Han population. *Neuroreport* **24**, 464-468.

Uncorrected Author Proof

Summary

In this work, we have devised a reproducible data-and-knowledge integration framework to identify pleiotropic genes and their roles in AD and T2DM. By identifying important shared genetic variants in AD and T2DM, followed by meta-analysis of omics data, we have performed a systematic functional assessment of these signals using a literature-derived knowledge assembly. This design overcomes the issue of publication bias because the frequency of recurrent studies does not influence the outcome of the data analysis. In other words, this study is purely guided by findings from analyses of data, while we have made use of literature to illustrate the functional roles of prioritized entities in AD-T2DM comorbidity. In doing so, we were able to prioritize four genes (i.e., COMT, MMP9, SOD2 and ABCG1) which were pleiotropic in nature. The involvement of these genes in both AD and T2DM-related events suggests that downstream effects of these genes could be the mechanistic bridge between the diseases. Interestingly, although these genes are not widely associated with AD and T2DM, we were able to demonstrate, with the help of a mechanism-centric interpretation, that dysfunctional activities of these genes are crucial for AD-T2DM comorbidity.

Chapter 4

Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation amplitude (CMPA) algorithm

Introduction

The clinical readouts of progressive disorders such as AD and PD are observed to follow a common pattern of either increment or decrement over time. For instance, the levels of cerebrospinal fluid (CSF) biomarkers such as amyloid beta peptides, phosphorylated MAPT and Lewy bodies increase as the disease progresses from early to mild and from mild to late stage. Similarly, brain volumes of patients as measured by MRI scans show that the volumes decrease (i.e. brain atrophy) with increasing AD severity. The occurrence of these events is attributed to dysregulated activities of genes, caused by their abnormal expressions. The interactions of such genes trigger a cascade of events impairing downstream signaling pathways. The impaired signaling pathways involved in a disease event are referred to as a disease mechanism. Although several NDD-specific mechanisms have been formulated through Alzpathway [131], PD maps [132] and NeuroMMSig [133], it remains an open question whether mechanisms behave differently in different stages of the disease, in the way that clinical readouts do. In this chapter, we introduce a novel algorithm to address this question. The Candidate Mechanism Perturbation Amplitude (CMPA) algorithm enables the quantification of disease mechanisms using gene expression profiles. To our knowledge, our methodology is the first to quantify disease mechanisms at spatial and temporal resolution. As a case study, we applied the algorithm to two mechanisms, one each for AD and PD. Our findings demonstrate that mechanisms can be quantified and that they regulate with different intensities across brain regions and stages of disease, as influenced by differentially expressed genes participating in the corresponding mechanism.

Karki et al. BMC Bioinformatics (2019) 20:494
https://doi.org/10.1186/s12859-019-3101-1

BMC Bioinformatics

METHODOLOGY ARTICLE Open Access

Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation amplitude (CMPA) algorithm

Reagon Karki[1,2], Alpha Tom Kodamullil[1,2], Charles Tapley Hoyt[1,2] and Martin Hofmann-Apitius[1,2*]

Abstract

Background: Literature derived knowledge assemblies have been used as an effective way of representing biological phenomenon and understanding disease etiology in systems biology. These include canonical pathway databases such as KEGG, Reactome and WikiPathways and disease specific network inventories such as causal biological networks database, PD map and NeuroMMSig. The represented knowledge in these resources delineates qualitative information focusing mainly on the causal relationships between biological entities. Genes, the major constituents of knowledge representations, tend to express differentially in different conditions such as cell types, brain regions and disease stages. A classical approach of interpreting a knowledge assembly is to explore gene expression patterns of the individual genes. However, an approach that enables quantification of the overall impact of differentially expressed genes in the corresponding network is still lacking.

Results: Using the concept of heat diffusion, we have devised an algorithm that is able to calculate the magnitude of regulation of a biological network using expression datasets. We have demonstrated that molecular mechanisms specific to Alzheimer (AD) and Parkinson Disease (PD) regulate with different intensities across spatial and temporal resolutions. Our approach depicts that the mitochondrial dysfunction in PD is severe in cortex and advanced stages of PD patients. Similarly, we have shown that the intensity of aggregation of neurofibrillary tangles (NFTs) in AD increases as the disease progresses. This finding is in concordance with previous studies that explain the burden of NFTs in stages of AD.

Conclusions: This study is one of the first attempts that enable quantification of mechanisms represented as biological networks. We have been able to quantify the magnitude of regulation of a biological network and illustrate that the magnitudes are different across spatial and temporal resolution.

Keywords: Alzheimer's disease, Parkinson's disease, Mitochondrial dysfunction, Aggregation of neurofibrillary tangles, OpenBEL

Background

In recent years, systems biology approaches have played a pivotal role in the integration of multi-scale and multi-modal aspects of diseases. Knowledge assembly, one of the key outcomes of systems biology, connects entities such as genes, proteins, chemicals, miRNAs, genetic and epigenetic variants, biological processes, and phenotypes of a disease. These are represented as a set of biological networks with edges defining the types of relationships between the entities. Pathway databases such as KEGG [1], Reactome [2], and WikiPathways [3] have undertaken massive efforts of extracting and encoding biological information from the published literature to graphically depict complex biological networks as pathways. They serve as a repository of protein-protein interactions (PPIs), metabolic pathways, signal transduction pathways, cell-cell signaling pathways, and other cellular

* Correspondence: martin.hofmann-apitius@scai.fraunhofer.de
[1]Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany
[2]Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn-Aachen International Center for IT, Endenicher Allee 19a, 53115 Bonn, Germany

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

processes. They have been regarded as comprehensive knowledge assemblies for functional interpretation of genomics and provide information about characteristics, progression and aetiology of a disease. A total of 521, 2176, and 2677 pathways are represented in KEGG, Reactome, and WikiPathways respectively. These databases provide pathways in standard formats (e.g., Systems Biology Markup Language (SBML) [4] and Biological Pathway Exchange (BioPAX) [5]), enabling easy exchange of data and implementation into algorithms for visualization, simulation and analysis [6].

However, pathway databases do have some limitations. Firstly, they lack context specific representation of knowledge when it comes to disease specificity. Pathways are generalized representations of established cascade of events within a specific pathway boundary. For example, the insulin signaling pathway in KEGG draws from experimental evidence from different diseases including diabetes [7], cancer [8], and hamartoma syndrome [9]. Moreover, pathways are abstractions that have been delineated arbitrarily and do not necessarily represent pathophysiology processes (e.g., the crosstalk between insulin signaling pathway and neurotrophin signaling pathway) [10]. Secondly, the spectrum of biological information captured by pathways is limited. They are mostly populated with proteins, making them uni-modal content wise. They completely lack representation of biomarkers, genetic variations, epigenetics (genetic modifications), neuroimaging, and clinical features. For example, the Parkinson's disease (PD) network in KEGG does not include many significant entities which play a crucial role in PD, such as the methylation of KCNH1 [11], the rs393152 variant in CRHR1 [12], and S87 SNCA phosphorylation [13]. Moreover, the fact that the map has been developed by retrieving information from 20 scientific articles (with the latest citation from 2013) infers that it is not up-to-date and incomplete [14]. Lastly, pathways are neither species, tissue, nor cell type specific. The representations in pathway databases are derived from various organisms (e.g., human, mouse, rat, and drosophila) where each species is indicated by differently colored nodes. However, interactions at the molecular level in a pathway can differ in these conditions. A study by Seok et al. (2013) reported on poor recapitulation of genomic responses of human inflammatory diseases in mouse models [15]. Warren et al. (2015) reconfirmed essential differences between these two species at the molecular level by showing that mouse models mimicked only 12% of the genes dysregulated in human conditions [16]. These studies clearly suggest that entities involved in pathways can be specific to species, tissue, cell types, and especially diseases.

Lately, there have been a few independent studies suggesting that a disease-specific mechanism differ from the canonically represented pathways in KEGG or Reactome. Kodamullil et al. (2015) have illustrated two different mechanisms on how the neurotrophin signaling pathway is regulated under normal conditions and AD [17]. Furthermore, Karki et al. (2017) have mechanistically represented the crosstalk between the insulin signaling pathway and neurotrophin signaling pathway, explaining the underlying comorbid association between AD and Type 2 Diabetes Mellitus (T2DM) [10]. Disease specific knowledge representations have improved significantly over the years due to the advancement in resources, frameworks and aforementioned limitations in the pathway databases. Several frameworks such as SMBL, GeneMania, Malacards, and OpenBEL, developed with either pathway-centric or integrated molecular network or knowledge graph approaches, are capable of representing knowledge at extent of their own features and advantages [18]. Nevertheless, these frameworks share the drawback of lacking a strategy to rank and prioritize pathways and mechanisms (i.e., knowledge sub-graphs) with the existing pathway databases. The selection of important individual graphs is often influenced by literature bias or expert's opinion. A scoring schema that takes in to account measurable biological entities will enable researchers to overcome any biases and identify important mechanisms involved in a disease.

Several algorithms have been proposed to use pathway databases to assist in the interpretation of high-throughput *-omics* data. Drier et al. (2013) introduced the Pathifier algorithm to score dysregulated pathways in tumor samples [19]. While it is able to transform gene level information to pathway level information, it does not take into account the polarity of relationships (i.e. increase or decrease) between the genes involved. Catlett et al. (2013) devised Reverse Causal Reasoning (RCR), a reverse engineering method to detect mechanistic hypotheses from molecular profiling data that generates and scores hypothesis networks (HYPs) i.e., literature-derived causal networks consisting of an upstream node and its first downstream neighbors [20]. Similarly, Martin et al. (2014) proposed the Network Perturbation Amplitude (NPA) algorithm to assess HYPs using high-throughput measurement data and demonstrated its ability to quantify TNF-induced perturbation of inflammatory signaling [21]. Although the RCR and NPA algorithms consider both the expression levels of genes and the relationship types between genes in a network, they have the following limitations: 1) the applications are restricted to interpret treatment-induced and dose-dependent changes in activity, 2) the size of the network is too small as it only accounts for the first neighbors and 3) the interlink between HYPs (i.e. one HYP being regulator of another HYP) is not considered.

Molecular mechanisms associated with a disease are often complex; they contain cascade of events regulated

by biomolecules which collectively influence biological processes and signaling pathways. Therefore, considering disease mechanisms we should be able to quantify them beyond HYPs or a network with few levels of neighbors (i.e. first and second neighbors). In fact, several cross-linked HYPs can form a basis for larger networks representing models of pathological events or disease mechanisms. Therefore, it is of the utmost importance to extend interplays between entities from HYPs to biological process, biological process to pathways, and pathways to mechanisms. Additionally, as genes tend to express differentially in different bodily regions or stages of a disease, the mechanisms in which they participate can be upregulated or downregulated by combined effect of the differentially expressed entities. To address these limitations, we have developed an extension to the NPA algorithm which is able to quantify mechanisms by scoring all of their constituent entities. As a case study, we ran the algorithm over two mechanisms (i.e. mitochondrial dysfunction in PD and aggregation of neurofibrillary tangles (NFTs) in AD) after mapping with gene expression datasets. The main objective of the study is to find out if mechanisms are regulated with different intensities as a consequence of differentially expressed genes at several resolutions.

Results

In this study, we have deployed the CMPA algorithm on two mechanisms, one each from PD and AD. This has allowed us to quantify perturbed mechanisms and show that the amplitude of the perturbations are affected by the differentially expressed genes. Moreover, the algorithm is able to handle mechanistic information at spatial and temporal resolution.

Mitochondrial dysfunction in PD

The CMPA analysis of mitochondrial dysfunction in different age-groups of PD patients depicts that the mechanism is perturbed the most in age-group 40–50 when compared to other age-groups (Fig. 1a). The magnitude of perturbation calculated as CMPA score is 4.8. Supporting this result, Lesage et al. (2016) implicate the role of mitochondrial dysfunction in the early onset of PD. Similarly, Fig. 1b shows the highest perturbation of mitochondrial dysfunction in Braak 5–6 stage of PD patients with CMPA score of 4.9. In contrast, Braak Stages 1–2 and 3–4 show less perturbation or no perturbation with CMPA scores of 0.93 and 0.08, respectively. A study by Hattingen et al. (2009) supports the role of mitochondrial dysfunction in both early and advanced stages of PD. This shows that our results (Fig. 1a and b)

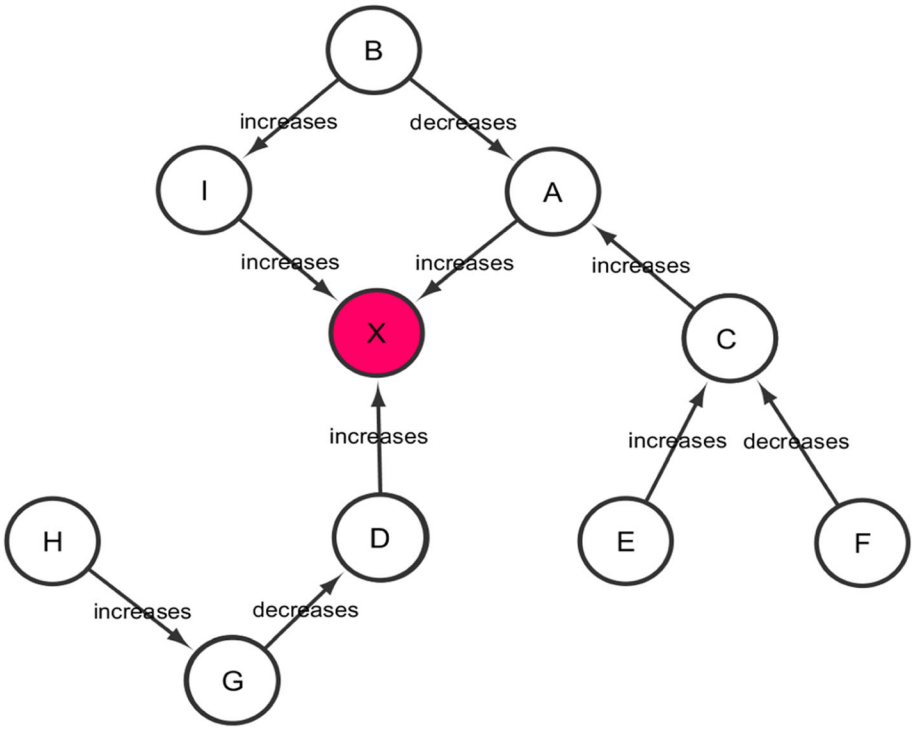

Fig. 1 Mechanisms perturb with different intensities: **a**, **b** and **c** show the amplitude of mitochondrial dysfunction in PD across age-groups, PD stages and brain regions respectively. The CMPA scores observed to be high in age-group 40–50, Braak Stage 5–6 and cortex of PD patients. Similarly, **d** shows the perturbation of aggregation of NFTs in AD across different stages of AD. The CMPA scores are observed to be directly proportional with stages of AD

are in concordance with other independent studies performed at the patient level. Interestingly, it can be seen in Fig. 1a and b that the amplitude of perturbation is low in age-group 50–60, 60–70, and Braak Stage 3–4. The rationale for these observations may be due to immunity triggered recovery or/and effect of drug used for treatment of PD. The inefficacy of both the immune system and the drug might be the reason for increased mitochondrial dysfunction in Braak 5–6 Stage of PD. Furthermore, Fig. 1c illustrates that the degree of perturbation of mitochondrial function varies across brain regions of PD patients. With CMPA score of 3.3, cortex is the region of the brain with the highest mitochondrial dysfunction. The magnitude of dysfunctions in other brain regions such as the cerebellum, medulla and striatum are minimal in comparison [22]. In this context, several animal and human based studies have previously confirmed prevalence of mitochondrial dysfunction in cortex [23–25].

Aggregation of NFTs in AD

The CMPA scores calculated for different stages of AD as shown in Fig. 1d suggests that the intensity with which aggregation of NFTs is regulated depends upon the stage of AD. The CMPA scores of incipient, moderate and severe AD are 3.6, 8.2 and 16.5 respectively. It can be clearly observed that the CMPA scores are directly correlated with the stages of AD. This comprehensively alienates with the findings of increased NFT burden with the progression of AD as reported by several studies [26–28].

Discussion

As the NeuroMMSig server embeds numerous molecular signatures implicated in AD and PD, it provides us the opportunity to extend the CMPA analysis beyond the two mechanisms we have undertaken in this study. An extensive implementation of the CMPA algorithm on NeuroMMsig based mechanisms will enable us to rank mechanisms based on the CMPA scores. By scoring mechanisms on several resolutions, we may be able to prioritize the targetable mechanisms and thereby decide on the best suited medicine. For example, the CMPA score of 0.08 for mitochondrial dysfunction in a PD patient of Braak Stage 3–4 suggests reduced perturbation of the mechanism. Hence, targeting dysfunctional mitochondrial activity for patients with Braak 3–4 stage of PD might not be as important as it is for Braak 5–6 stage of PD. This sort of approach defies any literature bias, where one mechanism can be overly represented in a knowledge network because of the high density of supporting publications.

CMPA scores are mechanism specific

It has been observed that the CMPA scores are unique for all the gene expression datasets used in this study. Therefore, for each sub-groups of these datasets we have essentially been able to show that mechanisms are regulated with different magnitudes. The one sample t-test for GSE57475's age-group 40–50 in PD rejected the null hypothesis with a p-value < 2.2e-16 and t-statistic of − 166. The mean of 10,000 CMPA scores was 0.19 as compared to the actual CMPA score of 4.8. Similarly, the null hypothesis for GSE28146's moderate sub-group of AD was also rejected as the mean of CMPA scores and actual CMPA score were 1.77 and 8.2 respectively. Therefore, the alternative hypothesis i.e., true mean is not equal to 8.2 was favored with a p-value < 2.2e-16 and t-statistic of − 67.19. These results suggest that the CMPA score obtained from the real gene expression values is unique to a mechanism and is highly unlikely to occur just by chance.

Conclusions

In this study, we have demonstrated that blending computable knowledge and data in a given disease context provides us with new options for inference. Although strategies to integrate knowledge driven and data driven approaches already exist, our work deals with two new aspects: Firstly, we have been able to quantify candidate mechanisms underlying diseases. This is novel when compared to previous studies because we claim that our work is one of the first attempts to score complex biological networks that explain disease etiology. The causal relationship in OpenBEL, which forms the basis of making the OpenBEL knowledgebase computable, is the key in devising the CMPA algorithm. Without the information on the causality of the interacting biological entities, measuring the amplitude of a regulated mechanism is not possible. Secondly, we could demonstrate that differentially expressed genes regulate their corresponding mechanisms with different intensities. The differences in regulation intensities of mechanisms in temporal and spatial resolution have been reported through our study for the very first time. Based on the CMPA algorithm applied on 3 selected GE datasets, we observed that PD patients of Braak Stage 5–6, the age-group 40–50 and the cortex region of the brain have high magnitudes of mechanism perturbation. Similarly, we found out that the magnitudes of perturbation of aggregation of NFTs in AD increase with the progression of AD. From our results, we can conclude that the classical approach of associating mechanisms to progressive disorders can be improved by quantifying and prioritizing specifics such as disease stages, patient groups and brain regions.

Methods

Construction of mechanistic NDD knowledgebase

The unstructured textual information containing cause-and-effect or correlative relationships from literature specific to AD and PD were encoded as triples (i.e. subject-predicate-object) using OpenBEL. Furthermore, the triples are enriched with meta-annotations such as cell type, species, anatomy and stage of the disease. With additional curation efforts, each triple was assigned to a particular mechanistic sub-graph as described by Domingo-Fernandez et al. (2017) [29]. The resulting sub-graph contains several inter-connected triples depicting a disease mechanism. A total of 124 and 65 molecular mechanisms specific to AD and PD respectively are integrated in NeuroMMSig. For our analysis, we have taken into consideration the mechanisms depicting aggregation of neurofibrillary tangles (NFTs) in AD and mitochondrial dysfunction in PD. The mitochondrial dysfunction in PD is considered as one of the most important mechanisms associated with the PD etiology. Moreover, the AETIONOMY project (www.imi.europa.eu/projects-results/project-factsheets/aetionomy) has selected this mechanism for its intensive research. Similarly, the aggregation of NFTs in AD is a well-known AD phenotype and regarded as an important hypothesis in AD etiology. After filtering the mechanisms for causal relationships manually and using a threshold of five nearest neighbors as network size, the mechanism representing aggregation of NFTs in AD had a total of 31 nodes and 57 edges while the mitochondrial dysfunction in PD had 35 nodes and 54 edges (Additional file 1).

Selection of datasets as a scoring input

This study aims to quantify the intensity of perturbed mechanisms associated with diseases as the consequence of differentially expressed genes. Therefore, the candidate mechanism perturbation amplitude (CMPA) algorithm reduces the existing caveat of mere mechanism-disease associations by showing that mechanisms regulate with different intensities across spatial and temporal dimensions. Gene expression datasets from GEO (Gene Expression Omnibus) were selected such that the expression profiles could be categorized based on spatial dimensions (i.e., brain regions), temporal dynamics (i.e., age groups) or stages of the disease. These datasets were analyzed using GEO2R from GEO. A brief description of each of the datasets is given below:

I. **GSE49036** - Samples from Substantia nigra of different Braak Stages PD patients
II. **GSE57475** - Blood transcripts of PD patients of 4 different age groups
III. **GSE28894** - Samples from cerebellum, medulla, cortex, and striatum of PD patients
IV. **GSE28146** - Samples from Hippocampus of different stages of AD patients

Implementation of candidate mechanism perturbation amplitude (CMPA) algorithm

The strategy involved in this study is to integrate knowledge driven approaches and data driven approaches to score biological networks. Here, we have used gene expression profiles mapped to NeuroMMSig based causal networks to calculate the extent of perturbation of mechanism associated with AD and PD. A total of 3 datasets (i.e., GSE49036, GSE57475 and GSE28894) were mapped to the causal network representing mitochondrial dysfunction in PD while GSE28146 was mapped to the network representing aggregation of NFTs in AD. The causality between biological entities captured in BEL is one of the special features of BEL which many of the pathway representations are void of. Without the information about causal edges in disease networks, devising a scoring algorithm is not possible.

Scoring function

The expression profiles (i.e., log fold change values) are assigned as weights to the genes involved in a mechanism. The directionality of edges is taken from the mechanistic causal network as + 1 for increase and – 1 for decrease. A scoring function implemented in Python uses the weights and directionality of edges to quantify the amplitude of dysregulated mechanisms. A positive score implies that a particular mechanism for a given dataset is upregulated (i.e., perturbed) due to the interplay of involved downstream entities. Likewise, a negative score indicates that the mechanism is downregulated while a score of zero suggests no change in the mechanism.

Perturbation amplitude

The amplitude of perturbation is calculated for the central node (most upstream node) in the network to which several downstream nodes are connected. These downstream nodes can be either direct or indirect neighbors of the central node. Moreover, a downstream node can be a child node for other upstream nodes. Figure 2 illustrates a general cause-and-effect mechanism where downstream nodes converge to the centrally located node (node X, highlighted in red). The final score of the central node is calculated by enumerating the effect of differentially expressed downstream nodes on a particular mechanism context (in this case, the central node and the scored downstream nodes).

After this, the nodes outgoing from the central node were not considered (filtered and removed) as the central node mostly connects only either to another hub of the knowledgebase (in our case: Parkinson Disease) or to

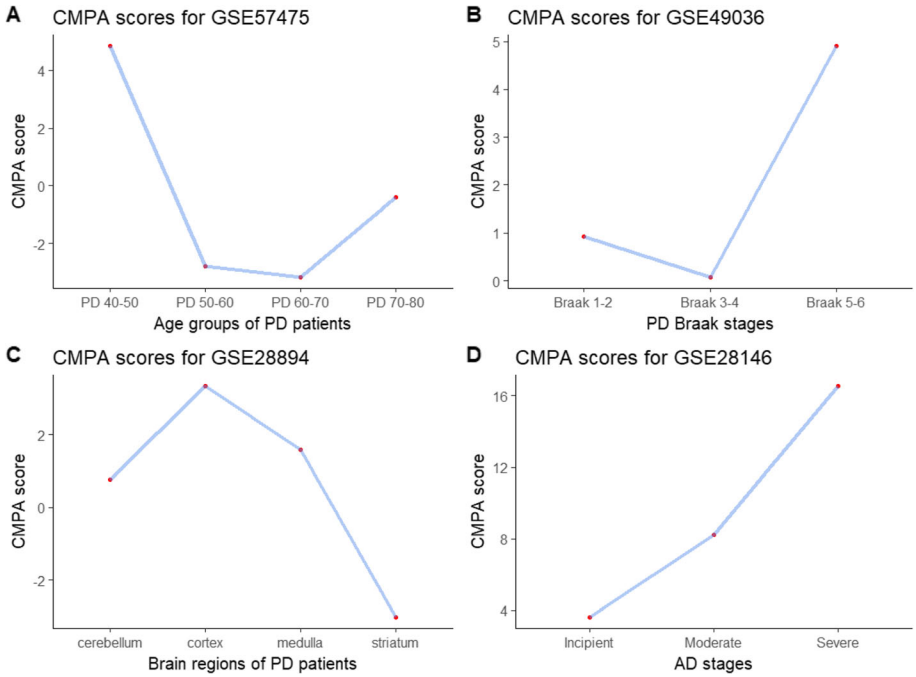

Fig. 2 A general biological network: A schematic representation of a mechanism where several upstream nodes (either genes/proteins or biological processes) converge to a centrally located node X (highlighted in red)

another central node (which can be another mechanism) and need not be scored.

The following *pseudocode* implemented in python was used to calculate the perturbation (CMPA) scores.

- Identify and create a list of hubs (H) in the network i.e. a node with several incoming and outgoing edges
- For each hub in H
  - If hub **has incoming** edges from another hub from the list H
    - Skip
  - If hub has **no incoming** edges from another hub from list H
    - Calculate Impact Factor (IF)
      IF = hubWeight + $\sum_{i=1}^{N} S_i \cdot \beta_i$
      where,
      S_i = Sign of the edge (+ 1 for increase, − 1 for decrease)
      β_i = Log_2 fold change value
      N = number of incoming nodes
    - Remove hub from H
- Calculate CMPA score
  - CMPA score = $\sum_{j=1}^{M} IF_i$
    Where,
    M = number of hubs

The CMPA algorithm is devised such that it is able to quantify the overall effect of differentially expressed entities involved in a cause-and-effect model of a disease mechanism. The algorithm functions on a simple logic that downstream nodes pass their values to the connected upstream nodes. For example, the value of H is passed to X through H − G − D − X (Fig. 1). In doing so, it is assured that G gets a value from H before G passes its value to D. The nodes G and D are hub nodes in the network because they have incoming and outgoing edges. For each hub node in the network, a score called Impact Factor (IF) is calculated. The sum of all the IFs, represented as CMPA score, quantify the amplitude of perturbation of a mechanism.

Statistical assessment of CMPA scores The CMPA scores generated by the CMPA algorithm are expected to be unique for each gene expression dataset. This is because of the distinct property of each gene responding differently to different conditions. However, a CMPA score can be considered absurd if it remains unchanged after random sampling of genes and their expressions. In the case differences in CMPA scores are observed between CMPA analyses performed with actual gene expressions and randomized gene expressions, it can be concluded that the CMPA score is specific to a mechanism and represents the true magnitude of its perturbation. This was assessed by first performing a permutation (number of permutations = 10,000) where each gene was assigned a random gene expression value from the pool of

real gene expression values. Afterwards, the CMPA algorithm was implemented to each of the permuted samples. Lastly, one sample Student's t-test was conducted with the null hypothesis that the mean of 10,000 CMPA scores is equal to the actual CMPA score. If the resulting *p*-value is below the threshold of 0.05, then the null hypothesis is rejected in favor of the alternative hypothesis.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12859-019-3101-1.

Additional file 1: Figure S1. Mitochondrial dysfunction in PD manifests as a consequence of increased oxidative stress and endoplasmic reticulum stress and decreased regulation of mitophagy. **Figure S2.** The aggregation of NFTs in AD is triggered by the insulin receptor signaling pathway and several genes that destabilize MAPT activity.

Abbreviations

AD: Alzheimer's Disease; BioPAX: Biological Pathway Exchange; CMPA: Candidate Mechanism Perturbation Amplitude; GEO: Gene Expression Omnibus; NFT: Neurofibrillary Tangle; NPA: Network Perturbation Amplitude; PD: Parkinson's Disease; RCR: Reverse Causal Reasoning; SMBL: Systems Biology Markup Language

Acknowledgements

Authors would like to thank scientific assistants who participated in curation of BEL statements. Lastly, authors would like to thank Sarah Mubeen for reviewing the manuscript and her consent to be acknowledged.

Authors' contributions

RK designed and implemented the devised algorithm. RK analyzed the expression datasets. RK and ATK built and analyzed the networks. RK wrote the manuscript. MHA, ATK and CTH contributed in writing the manuscript. CTH and MHA reviewed the manuscript. MHA supervised the study. All authors have read and approved the manuscript.

Funding

This work received financial support from the B-IT foundation that sponsors part of the academic work in our department. Part of this research has also received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115568 (project AETIONOMY), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2017–2013) and European Federation of Pharmaceutical Industries and Associations companies. The funding bodies did not play any role in design of the study and collection, analysis and interpretation of data and in writing the manuscript.

Availability of data and materials

1. GSE57475 - https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57475
2. GSE49036 - https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49036
3. GSE28894 - https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28894
4. GSE28146 - https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28146

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 7 May 2019 Accepted: 16 September 2019
Published online: 11 October 2019

References

1. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. {K}{E}{G}{G}: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45:D353–61.
2. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L. The reactome pathway knowledgebase. Nucleic Acids Res. 2015;44(D1):D481-7.
3. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Mélius J, Cirillo E, Coort SL, Digles D, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res. 2017;46:D661–7.
4. Hucka M, Bergmann FT, Drager A, Hoops S, Keating SM, Le Novere N, Myers CJ, Olivier BG, Sahle S, Schaff JC, Smith LP, Waltemath D, Wilkinson DJ. The Systems Biology Markup Language (SBML): language specification for level 3 version 2 core. J Integr Bioinform. 2018;15:1.
5. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, Schacherer F, Martinez-Flores I, Hu Z, Jimenez-Jacinto V, Joshi-Tope G, Kandasamy K, Lopez-Fuentes AC, Mi H, Pichler E, Rodchenkov I, Splendiani A, Tkachev S, Zucker J, Gopinath G, Rajasimha H, Ramakrishnan R, Shah I, Syed M, Anwar N, Babur O, Blinov M, Brauner E, Corwin D, Donaldson S, Gibbons F, Goldberg R, Hornbeck P, Luna A, Murray-Rust P, Neumann E, Ruebenacker O, Reubenacker O, Samwald M, van Iersel M, Wimalaratne S, Allen K, Braun B, Whirl-Carrillo M, Cheung KH, Dahlquist K, Finney A, Gillespie M, Glass E, Gong L, Haw R, Honig M, Hubaut O, Kane D, Krupa S, Kutmon M, Leonard J, Marks D, Merberg D, Petri V, Pico A, Ravenscroft D, Ren L, Shah N, Sunshine M, Tang R, Whaley R, Letovksy S, Buetow KH, Rzhetsky A, Schachter V, Sobral BS, Dogrusoz U, McWeeney S, Aladjem M, Birney E, Collado-Vides J, Goto S, Hucka M, Le Novere N, Maltsev N, Pandey A, Thomas P, Wingender E, Karp PD, Sander C, Bader GD. {T} he {B} io {P}{A}{X} community standard for pathway data sharing. Nat Biotechnol. 2010;28:935–42.
6. Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. Mol Syst Biol. 2009;5:290.
7. Ogawa W. Mechanism of insulin action and diabetes mellitus. Seiakgaku. 2003;75:1332–44.
8. Ruggero D, Sonenberg N. {T} he {a} kt of translational control. Oncogene. 2005;24:7426–34.
9. Inoki K, Corradetti MN, Guan KL. {D} ysregulation of the {T}{S}{C}-m {T}{O}{R} pathway in human disease. Nat Genet. 2005;37:19–24.
10. Karki R, Kodamullil AT, Hofmann-Apitius M. Comorbidity analysis between Alzheimer's disease and type 2 diabetes mellitus (T2DM) based on shared pathways and the role of T2DM drugs. J Alzheimers Dis. 2017;60:721–31.
11. Masliah E, Dumaop W, Galasko D, Desplats P. {D} istinctive patterns of {D}{N}{a} methylation associated with {P} arkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes. Epigenetics. 2013;8:1030–8.
12. Desikan RS, Schork AJ, Wang Y, Witoelar A, Sharma M, LK ME, Holland D, Brewer JB, Chen CH, Thompson WK, Harold D, Williams J, Owen MJ, O'Donovan MC, Pericak-Vance MA, Mayeux R, Haines JL, Farrer LA, Schellenberg GD, Heutink P, Singleton AB, Brice A, Wood NW, Hardy J, Martinez M, Choi SH, DeStefano A, Ikram MA, Bis JC, Smith A, Fitzpatrick AL, Launer L, van Duijn C, Seshadri S, Ulstein ID, Aarsland D, Fladby T, Djurovic S, Hyman BT, Snaedal J, Stefansson H, Stefansson K, Gasser T, Andreassen OA, Dale AM. {G} enetic overlap between {A}lzheimer's disease and {P}arkinson's disease at the {M}{A}{P}{T} locus. Mol Psychiatry. 2015;20:1588–95.
13. Taymans JM, Baekelandt V. {P} hosphatases of Î±-synuclein, {L}{R}{R}{K}2, and tau: important players in the phosphorylation-dependent pathology of {P}arkinsonism. Front Genet. 2014;5:382.
14. Wadi L, Meyer M, Weiser J, Stein LD, Reimand J. {I} mpact of outdated gene annotations on pathway enrichment analysis. Nat Methods. 2016;13:705–6.
15. Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, Richards DR, GP MD-S, Gao H, Hennessy L, Finnerty CC, Lopez CM, Honari S, Moore EE, Minei JP, Cuschieri J, Bankey PE, Johnson JL, Sperry J, Nathens AB, Billiar TR, West MA, Jeschke MG, Klein MB, Gamelli RL, Gibran NS, Brownstein BH, Miller-Graziano C, Calvano SE, Mason PH, Cobb JP, Rahme LG, Lowry SF, Maier RV, Moldawer LL, Herndon DN, Davis RW, Xiao W, Tompkins RG, Abouhamze A, Balis UG, Camp DG, De AK HBG, Hayden DL, Kaushal A,

O'Keefe GE, Kotz KT, Qian W, Schoenfeld DA, Shapiro MB, Silver GM, Smith RD, Storey JD, Tibshirani R, Toner M, Wilhelmy J, Wispelwey B, Wong WH. {G} enomic responses in mouse models poorly mimic human inflammatory diseases. Proc Natl Acad Sci U S A. 2013;110:3507–12.
16. Warren HS, Tompkins RG, Moldawer LL, Seok J, Xu W, Mindrinos MN, Maier RV, Xiao W, Davis RW. {M} ice are not men. Proc Natl Acad Sci U S A. 2015; 112:E345.
17. Kodamullil AT, Younesi E, Naz M, Bagewadi S, Hofmann-Apitius M. Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis. Alzheimers Dement. 2015; 11(11):1329-39.
18. Saqi M, Lysenko A, Guo YK, Tsunoda T, Auffray C. Navigating the disease landscape: knowledge representations for contextualizing molecular signatures. Brief Bioinform. 2018;20(2):609-23.
19. Drier Y, Sheffer M, Domany E. {P}athway-based personalized analysis of cancer. Proc Natl Acad Sci U S A. 2013;110:6388–93.
20. Catlett NL, Bargnesi AJ, Ungerer S, Seagaran T, Ladd W, Elliston KO, Pratt D. {R} everse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data. BMC Bioinformatics. 2013;14:340.
21. Martin F, Thomson TM, Sewer A, Drubin DA, Mathis C, Weisensee D, Pratt D, Hoeng J, Peitsch MC (2012) {A} ssessment of network perturbation amplitudes by applying high-throughput data to causal biological networks. BMC Syst Biol 6, 54.
22. Navarro A, Boveris A. Brain mitochondrial dysfunction in aging, neurodegeneration, and Parkinson's disease. Front Aging Neurosci. 2010;2:34.
23. Ferrer I. {E} arly involvement of the cerebral cortex in {P}arkinson's disease: convergence of multiple metabolic defects. Prog Neurobiol. 2009;88:89–103.
24. Stichel CC, Zhu XR, Bader V, Linnartz B, Schmidt S, Lubbert H. {M}ono- and double-mutant mouse models of {P}arkinson's disease display severe mitochondrial damage. Hum Mol Genet. 2007;16:2377–93.
25. Gautier CA, Kitada T, Shen J. {L} oss of {P}{I}{N}{K}1 causes mitochondrial functional defects and increased sensitivity to oxidative stress. Proc Natl Acad Sci U S A. 2008;105:11364–9.
26. Theofilas P, Ehrenberg AJ, Nguy A, Thackrey JM, Dunlop S, Mejia MB, Alho AT, Paraizo Leite RE, Rodriguez RD, Suemoto CK, Nascimento CF, Chin M, Medina-Cleghorn D, Cuervo AM, Arkin M, Seeley WW, Miller BL, Nitrini R, Pasqualucci CA, Filho WJ, Rueb U, Neuhaus J, Heinsen H, Grinberg LT. {P} robing the correlation of neuronal loss, neurofibrillary tangles, and cell death markers across the {a}lzheimer's disease {B} raak stages: a quantitative study in humans. Neurobiol Aging. 2018;61:1–12.
27. Braak H, Braak E. {S} taging of {A}lzheimer's disease-related neurofibrillary changes. Neurobiol Aging. 1995;16:271–8.
28. Jeong S. {M} olecular and {C} ellular {B} asis of {N} eurodegeneration in {a}lzheimer's {D}isease. Mol Cells. 2017;40:613–20.
29. Domingo-Fernández D, Kodamullil AT, Iyappan A, Naz M, Emon MA, Raschka T, Karki R, Springstubbe S, Ebeling C, Hofmann-Apitius M. Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): a web server for mechanism enrichment. Bioinformatics. 2017;33:3679–81.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

Summary

With our methodology of integrating omics data with knowledge assembly-derived mechanisms, we have successfully scored and quantified disease mechanisms. Our approach thus provides causal explanations, rather than mere association, for the relationships between mechanisms and diseases. Our finding that the CMPA scores for aggregation of NFTs positively correlates with stages of AD is in concordance with reports from previous studies reporting increased levels of NFTs with progression of AD. This proof of concept of our methodology justifies its implementation to several other mechanisms to generate a ranked list of important mechanisms. Similarly, we were also able to demonstrate in a PD mechanism, namely, mitochondrial dysfunction, that the intensity of its regulation varies in different regions of the brain. We identified that the intensity of regulation is highest in the cortex, which was also supported by findings from previous studies.

The causality of the interacting entities in the depicted mechanism forms the basis of the CMPA algorithm because it adheres to the principle of heat-diffusion such that the effects of differentially expressed genes are inherited by downstream biological processes which eventually lead to a disease mechanism. To the question whether CMPA scores are specific to a given mechanism, we were able to clarify the doubt by performing one sample t-test. We compared the mean of 10,000 randomly generated CMPA scores with the actual CMPA score. The alternative hypothesis (i.e., true mean is not equal to actual CMPA score) was favored with a statistically significant confidence ($p\text{-}value < 0.005$).

To sum up, our work has demonstrated that disease mechanisms can be quantified using omics data. Furthermore, we have also illustrated that the intensities of regulation of mechanisms 1) vary in disease stages and brain regions and 2) are directly influenced by expression patterns of corresponding genes that regulate the mechanism.

Chapter 5

Conclusion and Outlook

The principles of systems biology enhance the principles of classical biology, because systems biology invokes the need of achieving a holistic view of a system, as influenced by its components, in order to understand the behavior of the system. This has enabled us to create models that mimic biological systems and find out how information from granular level of genetic variants propagates to the higher levels of phenotypes and disease mechanisms. Furthermore, it has facilitated smooth communication between the ever-growing knowledge and data worlds. In fact, these domains are integral and complementary to each other. Despite the philosophical differences between classical and systems biology, the latter should not be seen as a replacement for the former. Rather, classical biology has evolved to produce systems biology, facilitated by advances in data collection and analysis and knowledge production. This thesis implements state-of-the-art technologies of systems biology to shed light on AD-T2DM comorbidity and formulate a new algorithm to quantify disease mechanisms.

In Chapter 2 and 3, we used two different methodologies to elucidate on the comorbid link between AD and T2DM as indicated by epidemiological, clinical and pathophysiological studies. In Chapter 2, we systematically retrieved relevant biological information about AD and T2DM from literature to create disease-specific knowledge assemblies. The molecular interactions and their downstream causal effects were structured as a mechanistic graph to explain how salient features of AD and T2DM unfold. Although previously published researches have already attempted to explain AD-T2DM comorbidity, the events associated with comorbidity were rather discrete and lacked mechanistic insights. Therefore, in this study, we have assimilated such important events, enriched them and provided an overview of comorbidity between AD and T2DM. We admit that the results in this work do reflect some degree of publication bias because

our approach in understanding comorbidity is based on literature-derived knowledge assemblies. Nonetheless, this work has identified previously unrecognized cross-talk between several signal pathways that are involved in AD and T2DM pathogenesis. The concordance of our findings derived from the knowledge assemblies were assessed by independently analyzed omics data. Moreover, in this work, we have discussed the putative harmful effects induced by Metformin, hypothesizing that Metformin is one of the risk factors in the comorbid link between AD and T2DM. In Chapter 3, we implemented a reverse workflow, using data as the prime impetus of the study. This allowed us to perform a publication-bias-free analysis, an issue we discussed above. From our results, we could identify novel pleiotropic genes and their genetic variants whose dysfunctional activities were implicated in the pathogenesis of AD and T2DM and the comorbidity between the diseases. Starting from collection of all reported deleterious genetic variants of AD and T2DM, we implemented a series of analyses, such as LD analysis and SNP prioritization, to identify important shared SNPs. To this end, we integrated an independent meta-analysis of omics data followed by functional contextualization of important genes through creation of knowledge assemblies. We believe our work is reproducible and the workflow can be readily used to study other comorbid conditions.

The previous attempts to score signaling pathways were bolstered by the implementation of the CMPA algorithm as discussed in Chapter 4. In this work, we were able, for the first time, to demonstrate that disease mechanisms regulate with different intensities as a consequence of differentially expressed genes. In other words, we were able to quantify and compare the severity of impairment of mechanisms across temporal and spatial resolution. This suggests the potential application of the CMPA algorithm for the purpose of generating a ranked list of most perturbed mechanisms. This will eventually empower us to classify important disease associated

mechanisms and rationalize on whether mechanisms embedding well-known genes of diseases truly represent disease etiologies.

The works described in Chapter 2 and 4 have contributed significantly to AETIONOMY (www.aetionomy.eu), a project aimed at deciphering a mechanism-based taxonomy of AD and PD. The published stories of these chapters are smaller instances of the knowledge assemblies mentioned in this thesis. In fact, the knowledge assemblies created for the studies contained in this thesis form the basis of a mechanism-enrichment server called NeuroMMSig [133] where interactions from the knowledge assemblies were systematically arranged and clustered into individual disease mechanisms. The work described in Chapter 3 is one of the first published research works in COMMITMENT (www.gesundheitsforschung-bmbf.de/commitment), an ongoing project aimed at improving treatment of patients with comorbidities. Using the workflow of this work, we aim to study comorbidity between Schizophrenia and Bipolar Disorder in the near future.

The forging of disease-specific knowledge assemblies is the most important contribution of this thesis. However, the coverage and content of our knowledge assemblies fail to keep up with the massively growing literature. One of the major limitations of our knowledge assemblies lies in the number of publications that are incorporated into them. When compared to the total number of published articles for each of the diseases studied in this thesis, we have been able to include just a handful of articles, suggesting that our knowledge assemblies are not complete. This is mainly because the knowledge assemblies are created solely through manual curation. Automated workflows should be developed to accelerate the entire process. Nonetheless, we are convinced that our curation work is of the highest quality and represents multi-modal aspects of diseases ranging from the genomic to the phenotypic level.

This thesis work illustrates the power of systems biology, by successfully integrating data and knowledge driven approaches. By leveraging these, we have broadened and brought in new insights to what was previously known about the comorbid link between AD and T2DM. Moreover, we have also demonstrated the possibility of quantifying mechanisms which can be useful in identifying important mechanisms in a given state of a disease.

Bibliography

[1] S. Kesić, “Systems biology, emergence and antireductionism,” *Saudi J. Biol. Sci.*, vol. 23, no. 5, pp. 584–591, 2016.

[2] A. T. Kodamullil *et al.*, “Trial watch: Tracing investment in drug development for Alzheimer disease.,” *Nat. Rev. Drug Discov.*, vol. 16, no. 12, p. 819, 2017.

[3] M. Drack, W. Apfalter, and D. Pouvreau, “On the making of a system theory of life: Paul A Weiss and Ludwig von Bertalanffy’s conceptual connection,” *Q. Rev. Biol.*, vol. 82, no. 4, pp. 349–373, 2007.

[4] M. D. Mesarovic, “General systems theory and biology--view of a theoretician,” *Gen. Syst. Theory Biol. Springer*, 1968.

[5] E. D. Green, J. D. Watson, and F. S. Collins, “Human Genome Project: Twenty-five years of big biology,” *Nature*, vol. 526, no. 7571, pp. 29–31, 2015.

[6] S. Khanna, D. Domingo-Fernández, A. Iyappan, M. A. Emon, M. Hofmann-Apitius, and H. Fröhlich, “Using multi-scale genetic, neuroimaging and clinical data for predicting Alzheimer’s Disease and reconstruction of relevant biological mechanisms,” *Sci. Rep.*, vol. 8, no. 1, p. 11173, Jul. 2018.

[7] A. Iyappan, R. Karki, S. Madan, E. Younesi, M. Hofmann-Apitius, and others, “Of mice and men: comparative analysis of neuro-inflammatory mechanisms in human and mouse using cause-and-effect models,” *J. Alzheimer’s Dis.*, vol. 59, no. 3, pp. 1045–1055, 2017.

[8] A. S. L. Hansen, R. M. Lennen, N. Sonnenschein, and M. J. Herrgård, “Systems biology solutions for biochemical production challenges,” *Current Opinion in Biotechnology*, vol.

45. Elsevier Ltd, pp. 85–91, 01-Jun-2017.

[9] C. Yu and F. Kobeissy, “Systems biology applications to decipher mechanisms and novel biomarkers in CNS trauma,” in *Brain neurotrauma: Molecular, neuropsychological, and rehabilitation aspects*, CRC Press/Taylor & Francis, 2015.

[10] L. Pray, D. A. Relman, E. R. Choffnes, and others, *The science and applications of synthetic and systems biology: workshop summary*. National Academies Press, 2011.

[11] J. Nielsen, “Systems biology of metabolism: a driver for developing personalized and precision medicine,” *Cell Metab.*, vol. 25, no. 3, pp. 572–579, 2017.

[12] T. D. Querec *et al.*, “Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans,” *Nat. Immunol.*, vol. 10, no. 1, p. 116, 2009.

[13] H. I. Nakaya *et al.*, “Systems biology of vaccination for seasonal influenza in humans,” *Nat. Immunol.*, vol. 12, no. 8, p. 786, 2011.

[14] O. T. Schubert and R. Aebersold, “Microbial Proteome Profiling and Systems Biology: Applications to Mycobacterium tuberculosis.,” *Adv. Exp. Med. Biol.*, vol. 883, pp. 235–54, 2015.

[15] J. B. Moore and M. E. Weeks, “Proteomics and systems biology: current and future applications in the nutritional sciences,” *Adv. Nutr.*, vol. 2, no. 4, pp. 355–364, 2011.

[16] A. Tinafar, K. Jaenes, and K. Pardee, “Synthetic Biology Goes Cell-Free,” *BMC Biol.*, vol. 17, no. 1, p. 64, 2019.

[17] P. Villoslada, L. Steinman, and S. E. Baranzini, “Systems biology and its application to the understanding of neurological diseases,” *Ann. Neurol. Off. J. Am. Neurol. Assoc. Child*

Neurol. Soc., vol. 65, no. 2, pp. 124–139, 2009.

[18] M. Altaf-Ul-Amin, F. M. Afendi, S. K. Kiboi, and S. Kanaya, "Systems biology in the context of big data and networks," *Biomed Res. Int.*, vol. 2014, 2014.

[19] B. A. Maron and J. A. Leopold, "Systems biology: an emerging strategy for discovering novel pathogenetic mechanisms that promote cardiovascular disease," *Glob. Cardiol. Sci. Pract.*, vol. 2016, no. 3, 2016.

[20] V. S. Bharadhwaj *et al.*, "CLEP: A Hybrid Data-and Knowledge-Driven Framework for Generating Patient Representations," *bioRxiv*, 2020.

[21] A. R. Feinstein, "The pre-therapeutic classification of co-morbidity in chronic disease," *J. Chronic Dis.*, vol. 23, no. 7, pp. 455–468, 1970.

[22] R. Gijsen, N. Hoeymans, F. G. Schellevis, D. Ruwaard, W. A. Satariano, and G. A. M. van den Bos, "Causes and consequences of comorbidity: a review," *J. Clin. Epidemiol.*, vol. 54, no. 7, pp. 661–674, 2001.

[23] M. Roca *et al.*, "Prevalence and comorbidity of common mental disorders in primary care," *J. Affect. Disord.*, vol. 119, no. 1–3, pp. 52–58, 2009.

[24] J. Sánchez-Valle *et al.*, "Interpreting molecular similarity between patients as a determinant of disease comorbidity relationships," *Nat. Commun.*, vol. 11, no. 1, pp. 1–13, 2020.

[25] Y. Ko, M. Cho, J.-S. Lee, and J. Kim, "Identification of disease comorbidity through hidden molecular mechanisms," *Sci. Rep.*, vol. 6, p. 39433, 2016.

[26] M. A. Sattar, A. A. Al-Sughyer, and R. Siboo, "Coexistence of rheumatoid arthritis,

ankylosing spondylitis and dermatomyositis in a patient with diabetes mellitus and the associated linked HLA antigens," *Rheumatology*, vol. 27, no. 2, pp. 146–149, 1988.

[27] C. T. Hoyt, D. Domingo-Fernández, N. Balzer, A. Güldenpfennig, and M. Hofmann-Apitius, "A systematic approach for identifying shared mechanisms in epilepsy and its comorbidities," *Database*, vol. 2018, 2018.

[28] J. B. Tomblin and K. L. Mueller, "How can the comorbidity with ADHD aid understanding of language and speech disorders?," *Top. Lang. Disord.*, vol. 32, no. 3, p. 198, 2012.

[29] M. Naz, E. Younesi, and M. Hofmann-Apitius, "Systematic Analysis of GWAS Data Reveals Genomic Hotspots for Shared Mechanisms between Neurodegenerative Diseases," *J Alzheimers Dis Park.*, vol. 7, no. 368, pp. 460–2161, 2017.

[30] J. Hardy and D. Allsop, "Amyloid deposition as the central event in the aetiology of Alzheimer's disease," *Trends Pharmacol. Sci.*, vol. 12, pp. 383–388, 1991.

[31] A. Mudher and S. Lovestone, "Alzheimer's disease--do tauists and baptists finally shake hands?," *Trends Neurosci.*, vol. 25, no. 1, pp. 22–26, 2002.

[32] P. T. Francis, A. M. Palmer, M. Snape, and G. K. Wilcock, "The cholinergic hypothesis of Alzheimer's disease: a review of progress," *J. Neurol. Neurosurg. Psychiatry*, vol. 66, no. 2, pp. 137–147, 1999.

[33] R. Deane and B. V Zlokovic, "Role of the blood-brain barrier in the pathogenesis of Alzheimer's disease," *Curr. Alzheimer Res.*, vol. 4, no. 2, pp. 191–197, 2007.

[34] A. R. Kamer, R. G. Craig, A. P. Dasanayake, M. Brys, L. Glodzik-Sobanska, and M. J. de

Leon, “Inflammation and Alzheimer’s disease: possible role of periodontal diseases,” *Alzheimer’s Dement.*, vol. 4, no. 4, pp. 242–250, 2008.

[35] F. Yin, H. Sancheti, I. Patil, and E. Cadenas, “Energy metabolism and inflammation in brain aging and Alzheimer’s disease,” *Free Radic. Biol. Med.*, vol. 100, pp. 108–122, 2016.

[36] H. Xu, D. I. Finkelstein, and P. A. Adlard, “Interactions of metals and Apolipoprotein E in Alzheimer’s disease,” *Front. Aging Neurosci.*, vol. 6, p. 121, 2014.

[37] S. C. Waring and R. N. Rosenberg, “Genome-wide association studies in Alzheimer disease,” *Arch. Neurol.*, vol. 65, no. 3, pp. 329–334, 2008.

[38] T. Jonsson *et al.*, “Variant of TREM2 associated with the risk of Alzheimer’s disease,” *N. Engl. J. Med.*, vol. 368, no. 2, pp. 107–116, 2013.

[39] K. Blennow, M. J. de Leon, and H. Zetterberg, “Alzheimer’s disease,” *Lancet*, vol. 368, no. 9533, pp. 387–403, Jul. 2006.

[40] J. K. Cataldo, J. J. Prochaska, and S. A. Glantz, “Cigarette smoking is a risk factor for Alzheimer’s Disease: an analysis controlling for tobacco industry affiliation,” *J. Alzheimer’s Dis.*, vol. 19, no. 2, pp. 465–480, 2010.

[41] S. S. M. Fernández and S. M. L. Ribeiro, “Nutrition and Alzheimer disease,” *Clin. Geriatr. Med.*, vol. 34, no. 4, pp. 677–697, 2018.

[42] P. V. Moulton and W. Yang, “Air pollution, oxidative stress, and Alzheimer’s disease,” *J. Environ. Public Health*, vol. 2012, 2012.

[43] T. Vos *et al.*, “Global, regional, and national incidence, prevalence, and years lived with

disability for 310 diseases and injuries, 1990--2015: a systematic analysis for the Global Burden of Disease Study 2015," *Lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.

[44] S. Brunton and others, "Pathophysiology of type 2 diabetes: the evolution of our understanding," *J Fam Pr.*, vol. 65, no. 4 Suppl, p. supp_az_0416, 2016.

[45] M. S. German, "Chapter 17. Pancreatic Hormones and Diabetes Mellitus. Dalam: Gardner DG, Shoback D, eds. Greenspan's Basic & Clinical Endocrinology." New York: McGraw-Hill, 2011.

[46] A. Pan, Y. Wang, M. Talaei, F. B. Hu, and T. Wu, "Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis," *Lancet Diabetes Endocrinol*, vol. 3, no. 12, pp. 958–967, Dec. 2015.

[47] D. Mozaffarian, A. Kamineni, M. Carnethon, L. Djouss?, K. J. Mukamal, and D. Siscovick, "Lifestyle risk factors and new-onset diabetes mellitus in older adults: the cardiovascular health study," *Arch. Intern. Med.*, vol. 169, no. 8, pp. 798–807, Apr. 2009.

[48] V. S. Malik, B. M. Popkin, G. A. Bray, J. P. Despr?s, and F. B. Hu, "Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk," *Circulation*, vol. 121, no. 11, pp. 1356–1364, Mar. 2010.

[49] H. Izzedine, V. Launay-Vacher, C. Deybach, E. Bourry, B. Barrou, and G. Deray, "Drug-induced diabetes mellitus," *Expert Opin Drug Saf*, vol. 4, no. 6, pp. 1097–1109, Nov. 2005.

[50] U. K. Sampson, M. F. Linton, and S. Fazio, "Are statins diabetogenic?," *Curr. Opin. Cardiol.*, vol. 26, no. 4, pp. 342–347, Jul. 2011.

[51] R. Larsen and H. Kronenberg, “et ad. Williams textbook of endocrinology 12th edition.” Elsevier Saunders, 2011.

[52] C. Fuchsberger *et al.*, “The genetic architecture of type 2 diabetes,” *Nature*, vol. 536, no. 7614, pp. 41–47, 2016.

[53] C. Herder and M. Roden, “Genetics of type 2 diabetes: pathophysiologic and clinical relevance,” *Eur. J. Clin. Invest.*, vol. 41, no. 6, pp. 679–692, Jun. 2011.

[54] I. Moltke *et al.*, “A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes,” *Nature*, vol. 512, no. 7513, pp. 190–193, 2014.

[55] G. J. Cooper, A. C. Willis, A. Clark, R. C. Turner, R. B. Sim, and K. B. Reid, “Purification and characterization of a peptide from amyloid-rich pancreases of type 2 diabetic patients,” *Proc. Natl. Acad. Sci.*, vol. 84, no. 23, pp. 8628–8632, 1987.

[56] J. Leszek, E. Trypka, V. V Tarasov, G. Md Ashraf, and G. Aliev, “Type 3 diabetes mellitus: a novel implication of Alzheimers disease,” *Curr. Top. Med. Chem.*, vol. 17, no. 12, pp. 1331–1335, 2017.

[57] A. Hofman, D. E. Grobbee, P. De Jong, and F. A. den Ouweland, “Determinants of disease and disability in the elderly: the Rotterdam Elderly Study,” *Eur. J. Epidemiol.*, vol. 7, no. 4, pp. 403–422, 1991.

[58] M. Barbagallo and L. J. Dominguez, “Type 2 diabetes mellitus and Alzheimer’s disease,” *World J. Diabetes*, vol. 5, no. 6, p. 889, 2014.

[59] C. L. Leibson *et al.*, “The risk of dementia among persons with diabetes mellitus: a population-based cohort study,” *Ann. N. Y. Acad. Sci.*, vol. 826, pp. 422–427, Sep. 1997.

[60] C. C. Huang *et al.*, “Diabetes mellitus and the risk of Alzheimer’s disease: a nationwide population-based study,” *PLoS One*, vol. 9, no. 1, p. e87095, 2014.

[61] R. Peila, B. L. Rodriguez, and L. J. Launer, “Type 2 diabetes, APOE gene, and the risk for dementia and related pathologies: The Honolulu-Asia Aging Study,” *Diabetes*, vol. 51, no. 4, pp. 1256–1262, Apr. 2002.

[62] M. Takeda *et al.*, “Apolipoprotein E and central nervous system disorders: reviews of clinical findings,” *Psychiatry Clin. Neurosci.*, vol. 64, no. 6, pp. 592–607, Dec. 2010.

[63] G. A. Dore, M. F. Elias, M. A. Robbins, P. K. Elias, and Z. Nagy, “Presence of the APOE epsilon4 allele modifies the relationship between type 2 diabetes and cognitive performance: the Maine-Syracuse Study,” *Diabetologia*, vol. 52, no. 12, pp. 2551–2560, Dec. 2009.

[64] X. Li, D. Song, and S. X. Leng, “Link between type 2 diabetes and Alzheimer’s disease: from epidemiology to mechanism and treatment,” *Clin Interv Aging*, vol. 10, pp. 549–560, 2015.

[65] R. O. Roberts *et al.*, “Association of type 2 diabetes with brain atrophy and cognitive impairment,” *Neurology*, vol. 82, no. 13, pp. 1132–1141, Apr. 2014.

[66] L. A. Beckett, M. C. Donohue, C. Wang, P. Aisen, D. J. Harvey, and N. Saito, “The Alzheimer’s Disease Neuroimaging Initiative phase 2: Increasing the length, breadth, and depth of our understanding,” *Alzheimers Dement*, vol. 11, no. 7, pp. 823–831, Jul. 2015.

[67] C. Moran *et al.*, “Type 2 diabetes mellitus and biomarkers of neurodegeneration,” *Neurology*, vol. 85, no. 13, pp. 1123–1130, Sep. 2015.

[68] S. Chatterjee and A. Mudher, “Alzheimer’s Disease and Type 2 Diabetes: A Critical Assessment of the Shared Pathological Traits,” *Front Neurosci*, vol. 12, p. 383, 2018.

[69] N. Yamamoto, R. Ishikuro, M. Tanida, K. Suzuki, Y. Ikeda-Matsuo, and K. Sobue, “Insulin-signaling Pathway Regulates the Degradation of Amyloid β-protein via Astrocytes,” *Neuroscience*, vol. 385, pp. 227–236, 2018.

[70] S. O. Yoon *et al.*, “JNK3 perpetuates metabolic stress induced by Aβ peptides,” *Neuron*, vol. 75, no. 5, pp. 824–837, Sep. 2012.

[71] C. Benedict, M. Hallschmid, B. Schultes, J. Born, and W. Kern, “Intranasal insulin to improve memory function in humans,” *Neuroendocrinology*, vol. 86, no. 2, pp. 136–142, 2007.

[72] J. Freiherr *et al.*, “Intranasal insulin as a treatment for Alzheimer’s disease: a review of basic research and clinical evidence,” *CNS Drugs*, vol. 27, no. 7, pp. 505–514, Jul. 2013.

[73] W. Han and C. Li, “Linking type 2 diabetes and Alzheimer’s disease,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 15, pp. 6557–6558, Apr. 2010.

[74] J. J. Swaroop, D. Rajarajeswari, and J. N. Naidu, “Association of TNF-α with insulin resistance in type 2 diabetes mellitus.,” *Indian J. Med. Res.*, vol. 135, pp. 127–30, 2012.

[75] A. E. Hak *et al.*, “Markers of inflammation and cellular adhesion molecules in relation to insulin resistance in nondiabetic elderly: the Rotterdam study,” *J. Clin. Endocrinol. Metab.*, vol. 86, no. 9, pp. 4398–4405, Sep. 2001.

[76] G. S. Hotamisligil, “Inflammatory pathways and insulin action,” *Int. J. Obes. Relat. Metab. Disord.*, vol. 27 Suppl 3, pp. S53--55, Dec. 2003.

[77] S. H. Back and R. J. Kaufman, “Endoplasmic reticulum stress and type 2 diabetes,” *Annu. Rev. Biochem.*, vol. 81, pp. 767–793, 2012.

[78] A. Jaeschke, M. P. Czech, and R. J. Davis, “An essential role of the JIP1 scaffold protein for JNK activation in adipose tissue,” *Genes Dev.*, vol. 18, no. 16, pp. 1976–1980, Aug. 2004.

[79] C. Morel *et al.*, “Requirement of JIP1-mediated c-Jun N-terminal kinase activation for obesity-induced insulin resistance,” *Mol. Cell. Biol.*, vol. 30, no. 19, pp. 4616–4625, Oct. 2010.

[80] M. S. Bhamra and N. J. Ashton, “Finding a pathological diagnosis for Alzheimer’s disease: are inflammatory molecules the answer?,” *Electrophoresis*, vol. 33, no. 24, pp. 3598–3607, Dec. 2012.

[81] S. Sekar *et al.*, “Alzheimer’s disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes,” *Neurobiol. Aging*, vol. 36, no. 2, pp. 583–591, Feb. 2015.

[82] M. A. Fishel *et al.*, “Hyperinsulinemia provokes synchronous increases in central inflammation and beta-amyloid in normal adults,” *Arch. Neurol.*, vol. 62, no. 10, pp. 1539–1544, Oct. 2005.

[83] A. Sokolova, M. D. Hill, F. Rahimi, L. A. Warden, G. M. Halliday, and C. E. Shepherd, “Monocyte chemoattractant protein-1 plays a dominant role in the chronic inflammation observed in Alzheimer’s disease,” *Brain Pathol.*, vol. 19, no. 3, pp. 392–398, Jul. 2009.

[84] D. Doens and P. L. Fernández, “ Microglia receptors and their implications in the response to amyloid β for Alzheimer’s disease pathogenesis,” *J Neuroinflammation*, vol. 11, p. 48,

Mar. 2014.

[85] H. Akiyama *et al.*, “Inflammation and Alzheimer’s disease,” *Neurobiol. Aging*, vol. 21, no. 3, pp. 383–421, 2000.

[86] P. Eikelenboom, R. Veerhuis, W. Scheper, A. J. Rozemuller, W. A. van Gool, and J. J. Hoozemans, “The significance of neuroinflammation in understanding Alzheimer’s disease,” *J Neural Transm*, vol. 113, no. 11, pp. 1685–1695, Nov. 2006.

[87] H. Barron, S. Hafizi, A. C. Andreazza, and R. Mizrahi, “Neuroinflammation and Oxidative Stress in Psychosis and Psychosis Risk,” *Int J Mol Sci*, vol. 18, no. 3, Mar. 2017.

[88] M. Sheng, B. L. Sabatini, and T. C. Südhof, “Synapses and Alzheimer’s disease,” *Cold Spring Harb Perspect Biol*, vol. 4, no. 5, May 2012.

[89] C. A. Davies, D. M. Mann, P. Q. Sumpter, and P. O. Yates, “A quantitative morphometric analysis of the neuronal and synaptic content of the frontal and temporal cortex in patients with Alzheimer’s disease,” *J. Neurol. Sci.*, vol. 78, no. 2, pp. 151–164, Apr. 1987.

[90] F. M. LaFerla and S. Oddo, “Alzheimer’s disease: Abeta, tau and synaptic dysfunction,” *Trends Mol Med*, vol. 11, no. 4, pp. 170–176, Apr. 2005.

[91] G. M. Shankar and D. M. Walsh, “Alzheimer’s disease: synaptic dysfunction and Abeta,” *Mol Neurodegener*, vol. 4, p. 48, Nov. 2009.

[92] J. J. Palop and L. Mucke, “Amyloid-beta-induced neuronal dysfunction in Alzheimer’s disease: from synapses toward neural networks,” *Nat. Neurosci.*, vol. 13, no. 7, pp. 812–818, Jul. 2010.

[93] C. Haass and E. Mandelkow, “Fyn-tau-amyloid: a toxic triad,” *Cell*, vol. 142, no. 3, pp.

356–358, Aug. 2010.

[94] O. Katsuse, W. L. Lin, J. Lewis, M. L. Hutton, and D. W. Dickson, “Neurofibrillary tangle-related synaptic alterations of spinal motor neurons of P301L tau transgenic mice,” *Neurosci. Lett.*, vol. 409, no. 2, pp. 95–99, Dec. 2006.

[95] T. L. Spires-Jones and B. T. Hyman, “The intersection of amyloid beta and tau at synapses in Alzheimer’s disease,” *Neuron*, vol. 82, no. 4, pp. 756–771, May 2014.

[96] C. Tackenberg and R. Brandt, “Divergent pathways mediate spine alterations and cell death induced by amyloid-beta, wild-type tau, and R406W tau,” *J. Neurosci.*, vol. 29, no. 46, pp. 14439–14450, Nov. 2009.

[97] L. Jin *et al.*, “Cognitive deficits and Alzheimer-like neuropathological impairments during adolescence in a rat model of type 2 diabetes mellitus,” *Neural Regen Res*, vol. 13, no. 11, pp. 1995–2004, Nov. 2018.

[98] I. Granic, A. M. Dolga, I. M. Nijholt, G. van Dijk, and U. L. Eisel, “Inflammation and NF-kappaB in Alzheimer’s disease and diabetes,” *J. Alzheimers Dis.*, vol. 16, no. 4, pp. 809–821, 2009.

[99] A. Tumminia, F. Vinciguerra, M. Parisi, and L. Frittitta, “Type 2 Diabetes Mellitus and Alzheimer’s Disease: Role of Insulin Signalling and Therapeutic Implications,” *Int J Mol Sci*, vol. 19, no. 11, Oct. 2018.

[100] J. M. Gaspar, F. I. Baptista, M. P. Macedo, and A. F. Ambrósio, “Inside the Diabetic Brain: Role of Different Players Involved in Cognitive Decline,” *ACS Chem Neurosci*, vol. 7, no. 2, pp. 131–142, Feb. 2016.

[101] C. Balducci *et al.*, “Synthetic amyloid-beta oligomers impair long-term memory independently of cellular prion protein,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 5, pp. 2295–2300, Feb. 2010.

[102] S. Li, M. Jin, T. Koeglsperger, N. E. Shepardson, G. M. Shankar, and D. J. Selkoe, “Soluble Aβ oligomers inhibit long-term potentiation through a mechanism involving excessive activation of extrasynaptic NR2B-containing NMDA receptors,” *J. Neurosci.*, vol. 31, no. 18, pp. 6627–6638, May 2011.

[103] F. C. van Bussel *et al.*, “Increased GABA concentrations in type 2 diabetes mellitus are related to lower cognitive functioning,” *Med.*, vol. 95, no. 36, p. e4803, Sep. 2016.

[104] Y. J. Sheen and W. H. Sheu, “Association between hypoglycemia and dementia in patients with type 2 diabetes,” *Diabetes Res. Clin. Pr.*, vol. 116, pp. 279–287, Jun. 2016.

[105] N. Mizushima and M. Komatsu, “Autophagy: renovation of cells and tissues,” *Cell*, vol. 147, no. 4, pp. 728–741, Nov. 2011.

[106] R. A. Frake, T. Ricketts, F. M. Menzies, and D. C. Rubinsztein, “Autophagy and neurodegeneration,” *J. Clin. Invest.*, vol. 125, no. 1, pp. 65–74, Jan. 2015.

[107] M. V Blagosklonny, “TOR-centric view on insulin resistance and diabetic complications: perspective for endocrinologists and gerontologists,” *Cell Death Dis*, vol. 4, p. e964, Dec. 2013.

[108] R. A. Nixon, “Autophagy, amyloidogenesis and Alzheimer disease,” *J. Cell. Sci.*, vol. 120, no. Pt 23, pp. 4081–4091, Dec. 2007.

[109] K. Inoue *et al.*, “Macroautophagy deficiency mediates age-dependent neurodegeneration

through a phospho-tau pathway," *Mol Neurodegener*, vol. 7, p. 48, Sep. 2012.

[110] F. Pickford *et al.*, "The autophagy-related protein beclin 1 shows reduced expression in early Alzheimer disease and regulates amyloid beta accumulation in mice," *J. Clin. Invest.*, vol. 118, no. 6, pp. 2190–2199, Jun. 2008.

[111] W. Quan, Y. M. Lim, and M. S. Lee, "Role of autophagy in diabetes and endoplasmic reticulum stress of pancreatic β-cells," *Exp. Mol. Med.*, vol. 44, no. 2, pp. 81–88, Feb. 2012.

[112] Y. Fujitani, R. Kawamori, and H. Watada, "The role of autophagy in pancreatic beta-cell and diabetes," *Autophagy*, vol. 5, no. 2, pp. 280–282, Feb. 2009.

[113] Z. F. Chen *et al.*, "The double-edged effect of autophagy in pancreatic beta cells and diabetes," *Autophagy*, vol. 7, no. 1, pp. 12–16, Jan. 2011.

[114] C. Carvalho, M. S. Santos, C. R. Oliveira, and P. I. Moreira, "Alzheimer's disease and type 2 diabetes-related alterations in brain mitochondria, autophagy and synaptic markers," *Biochim. Biophys. Acta*, vol. 1852, no. 8, pp. 1665–1675, Aug. 2015.

[115] S. Grehan, E. Tse, and J. M. Taylor, "Two distal downstream enhancers direct expression of the human apolipoprotein E gene to astrocytes in the brain," *J. Neurosci.*, vol. 21, no. 3, pp. 812–822, Feb. 2001.

[116] R. W. Mahley, "Apolipoprotein E: cholesterol transport protein with expanding role in cell biology," *Science (80-.).*, vol. 240, no. 4852, pp. 622–630, Apr. 1988.

[117] V. Leduc, D. Domenger, L. De Beaumont, D. Lalonde, S. Bélanger-Jasmin, and J. Poirier, "Function and comorbidities of apolipoprotein e in Alzheimer's disease," *Int J Alzheimers*

Dis, vol. 2011, p. 974361, Apr. 2011.

[118] D. M. Hatters, N. Zhong, E. Rutenber, and K. H. Weisgraber, "Amino-terminal domain stability mediates apolipoprotein E aggregation into neurotoxic fibrils," *J. Mol. Biol.*, vol. 361, no. 5, pp. 932–944, Sep. 2006.

[119] J. Poirier, J. Davignon, D. Bouthillier, S. Kogan, P. Bertrand, and S. Gauthier, "Apolipoprotein E polymorphism and Alzheimer's disease," *Lancet*, vol. 342, no. 8873, pp. 697–699, Sep. 1993.

[120] K. Akter, E. A. Lanza, S. A. Martin, N. Myronyuk, M. Rua, and R. B. Raffa, "Diabetes mellitus and Alzheimer's disease: shared pathology and treatment?," *Br. J. Clin. Pharmacol.*, vol. 71, no. 3, pp. 365–376, 2011.

[121] T. Tokuda *et al.*, "Lipidation of apolipoprotein E influences its isoform-specific interaction with Alzheimer's amyloid beta peptides," *Biochem. J.*, vol. 348 Pt 2, pp. 359–365, Jun. 2000.

[122] I. J. Martins, T. Berger, M. J. Sharman, G. Verdile, S. J. Fuller, and R. N. Martins, "Cholesterol metabolism and transport in the pathogenesis of Alzheimer's disease," *J. Neurochem.*, vol. 111, no. 6, pp. 1275–1308, Dec. 2009.

[123] A. Fernandez-Gamba, M. C. Leal, L. Morelli, and E. M. Castano, "Insulin-degrading enzyme: structure-function relationship and its possible roles in health and disease," *Curr. Pharm. Des.*, vol. 15, no. 31, pp. 3644–3655, 2009.

[124] C. M. Carlsson, "Type 2 diabetes mellitus, dyslipidemia, and Alzheimer's disease," *J. Alzheimers Dis.*, vol. 20, no. 3, pp. 711–722, 2010.

[125] M. R. Taskinen, “Diabetic dyslipidemia,” *Atheroscler Suppl*, vol. 3, no. 1, pp. 47–51, May 2002.

[126] A. D. Mooradian, “Dyslipidemia in type 2 diabetes mellitus,” *Nat Clin Pr. Endocrinol Metab*, vol. 5, no. 3, pp. 150–159, Mar. 2009.

[127] Y. Tang, Y. M. Li, M. Zhang, Y. Q. Chen, and Q. Sun, “ε3/4 genotype of the apolipoprotein E is associated with higher risk of Alzheimer’s disease in patients with type 2 diabetes mellitus,” *Gene*, vol. 703, pp. 65–70, Jun. 2019.

[128] M. Malek-Ahmadi *et al.*, “Increased Alzheimer’s disease neuropathology is associated with type 2 diabetes and ApoE ε.4 carrier status,” *Curr Alzheimer Res*, vol. 10, no. 6, pp. 654–659, Jul. 2013.

[129] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, “KEGG: new perspectives on genomes, pathways, diseases and drugs,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D353--D361, 2017.

[130] A. Fabregat *et al.*, “The reactome pathway knowledgebase,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D649--D655, 2018.

[131] S. Ogishima *et al.*, “AlzPathway, an updated map of curated signaling pathways: towards deciphering Alzheimer’s disease pathogenesis,” in *Systems biology of Alzheimer’s disease*, Springer, 2016, pp. 423–432.

[132] K. A. Fujita *et al.*, “Integrating pathways of Parkinson’s disease in a molecular interaction map,” *Mol. Neurobiol.*, vol. 49, no. 1, pp. 88–102, 2014.

[133] D. Domingo-Fernández *et al.*, “Multimodal mechanistic signatures for neurodegenerative

diseases (NeuroMMSig): a web server for mechanism enrichment," *Bioinformatics*, vol. 33, no. 22, pp. 3679–3681, 2017.

[134] L. Gootjes-Dreesbach, M. Sood, A. Sahay, M. Hofmann-Apitius, and H. Fröhlich, "Variational Autoencoder Modular Bayesian Networks (VAMBN) for simulation of heterogeneous clinical study data," *BioRxiv*, p. 760744, 2019.

[135] Y. Drier, M. Sheffer, and E. Domany, "Pathway-based personalized analysis of cancer," *Proc. Natl. Acad. Sci.*, vol. 110, no. 16, pp. 6388–6393, 2013.

[136] D. Nishimura, "BioCarta," *Biotech Softw. Internet Rep. Comput. Softw. J. Sci.*, vol. 2, no. 3, pp. 117–120, 2001.

[137] C. F. Schaefer *et al.*, "PID: the pathway interaction database," *Nucleic Acids Res.*, vol. 37, no. suppl_1, pp. D674--D679, 2009.

[138] D. N. Slenter *et al.*, "WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D661--D667, 2018.

[139] D. Domingo-Fernández, C. T. Hoyt, C. Bobis-Álvarez, J. Mar\'\in-Llaó, and M. Hofmann-Apitius, "ComPath: an ecosystem for exploring, analyzing, and curating mappings across pathway databases," *NPJ Syst. Biol. Appl.*, vol. 4, no. 1, pp. 1–8, 2018.

[140] S. Mubeen, C. T. Hoyt, A. Gemünd, M. Hofmann-Apitius, H. Fröhlich, and D. Domingo-Fernández, "The impact of pathway database choice on statistical enrichment analysis and predictive modeling," *Front. Genet.*, vol. 10, p. 1203, 2019.

[141] L. Chindelevitch *et al.*, "Causal reasoning on biological networks: interpreting

transcriptional changes," *Bioinformatics*, vol. 28, no. 8, pp. 1114–1121, 2012.

[142] R. Kumar *et al.*, "Causal reasoning identifies mechanisms of sensitivity for a novel AKT kinase inhibitor, GSK690693," *BMC Genomics*, vol. 11, no. 1, pp. 1–12, 2010.

[143] N. L. Catlett *et al.*, "Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data," *BMC Bioinformatics*, vol. 14, no. 1, p. 340, 2013.

[144] F. Martin *et al.*, "Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks," *BMC Syst. Biol.*, vol. 6, no. 1, p. 54, 2012.

[145] A. Buniello *et al.*, "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1005--D1012, 2019.

[146] T. Beck, T. Shorter, and A. J. Brookes, "GWAS Central: a comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies," *Nucleic Acids Res.*, vol. 48, no. D1, pp. D933--D940, 2020.

[147] E. M. Smigielski, K. Sirotkin, M. Ward, and S. T. Sherry, "dbSNP: a database of single nucleotide polymorphisms," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 352–355, Jan. 2000.

[148] J. Piñero *et al.*, "The DisGeNET knowledge platform for disease genomics: 2019 update," *Nucleic Acids Res.*, vol. 48, no. D1, pp. D845–D855, 2019.

[149] I. A. Adzhubei *et al.*, "{A} method and server for predicting damaging missense mutations," *Nat. Methods*, vol. 7, no. 4, pp. 248–249, Apr. 2010.

[150] A. P. Boyle *et al.*, "Annotation of functional variation in personal genomes using

RegulomeDB," *Genome Res.*, vol. 22, no. 9, pp. 1790–1797, Sep. 2012.