Conceptualization of Computational Modeling Approaches and Interpretation of the Role of Neuroimaging Indices in Pathomechanisms for Pre-Clinical Detection of Alzheimer Disease

Kumulative Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Anandhi Iyappan

aus

Chennai, Indien

Bonn 2018

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. Thomas Schultz

Tag der Promotion: 10.09.2018
Erscheinungsjahr: 2018

Abstract

With swift advancements in next-generation sequencing technologies alongside of the voluminous growth of biological data, a diversity of various data resources such as databases and web services have been created to facilitate data management, accessibility, and analysis. However, the burden of interoperability between dynamically growing data resources is an increasingly rate-limiting step in biomedicine, specifically concerning neurodegeneration. Over the years, massive investments and technological advancements for dementia research have resulted in large proportions of unmined data. Accordingly, there is an essential need for intelligent as well as integrative approaches to mine available data and substantiate novel research outcomes. Semantic frameworks provide a unique possibility to integrate multiple heterogeneous, high-resolution data resources with semantic integrity using standardized ontologies and vocabularies for context-specific domains. In this current work, (i) the functionality of a semantically structured terminology for mining pathway relevant knowledge from the literature, called Pathway Terminology System, is demonstrated and (ii) a context-specific high granularity semantic framework for neurodegenerative diseases, known as NeuroRDF, is presented.

Neurodegenerative disorders are especially complex as they are characterized by widespread manifestations and the potential for dramatic alterations in disease progression over time. Early detection and prediction strategies through clinical pointers can provide promising solutions for effective treatment of AD. In the current work, we have presented the importance of bridging the gap between clinical and molecular biomarkers to effectively contribute to dementia research. Moreover, we address the need for a formalized framework called NIFT to automatically mine relevant clinical knowledge from the literature for substantiating high-resolution cause-and-effect models.

DECLARATION

I herewith declare that the present thesis is my original work, except where indicated through the proper use of citations and references. Any uses made within it of the works of other authors in any form are properly acknowledged at the point of their use. A full list of the references employed has been included.

Signature: ……..

Author: Anandhi Iyappan

Acknowledgement

First of all, I would like to express my deepest gratitude to Prof. Dr. Martin Hofmann-Apitius for giving me the opportunity to work on this thesis. He has provided with all degrees of freedom for me to grow from my comfort zone and to venture into new horizons. This has shaped me up into an independent scientist with the ability to conduct not just my own research but also guide bachelor and master theses students. His constant encouragement and immense support has been my pillar of strength.

I would also like to thank Prof. Dr. Thomas Schultz for instantly accepting to be my second supervisor. His infectious smile as well as constructive discussions was very helpful in writing my thesis.

I am eternally grateful to all my colleagues, students, and scientists at Fraunhofer SCAI for their friendly interaction, lively atmosphere and scientific support whenever I needed them. It would not be complete if I fail to mention my gratitude to Meike Knieps and Alina Enns for taking care of all the administrative support.

My special thanks to Dinesh Prabakaran, without whom my life in Germany would be incomplete. He is more than a friend, philosopher, guide and critic and has been the biggest pillar of strength and support to me all these years.

I cannot find words to express my love, gratitude and respect for my family members, who have sailed through all my ups and downs, who have dreamed my dreams, who have shouldered me when I needed them and also when I didn't need them.

I dedicate my thesis to my late grandmother Mrs. Meenakshi ammal, my late grandfather Subramaniam and to my dear mother Rajalakshmi Iyappan. She has been so selflessly thrown her life away to see me live my dreams; I am nothing without her in my life. I would forever be grateful to her.

Last but never the least, this is a small token of love to my eternal guru, Sadhguru. Words fail me to express my gratitude for him. I am forever grateful to have him in my life. He rightly says "Gratitude is something that flows out of you when you are overwhelmed with what has been given to you"

Publications

- **Iyappan, Anandhi**, et al. "Towards a pathway inventory of the human brain for modeling disease mechanisms underlying neurodegeneration." Journal of Alzheimer's Disease 52.4 (2016): 1343-1360.

- **Iyappan, Anandhi,** et al. "Neuroimaging Feature Terminology: A Controlled Terminology for the Annotation of Brain Imaging Features." Journal of Alzheimer's Disease (2017): 1153-1169.

- **Iyappan, Anandhi** et al. "Complexity across scales: a walkthrough to linking neuro-imaging readouts to molecular processes. " Journal of Systems and Integrative Neuroscience (2017).

- **Iyappan, Anandhi,** et al. NeuroRDF: semantic data integration strategies for modeling neurodegenerative diseases." Proceedings of the 6th International Symposium on Semantic Mining in Biomedicine (SMBM2014). 2014.

Other Publications:

- Kodamullil AT, **Iyappan, Anandhi**, 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 Preprint (2017): 1-11

- Domingo-Fernández, D, Kodamullil AT, **Iyappan, Anandhi**., Naz, M, Emon, M. A, Raschka, T, Karki, R, and Hofmann-Apitius, M. "Multimodal Mechanistic Signatures for Neurodegenerative Diseases (NeuroMMSig): a web server for mechanism enrichment." Bioinformatics (2017).

- Allen, GI, Amoroso, N, Anghel, C, Balagurusamy, V, Bare, CJ, Beaton, D**, Iyappan Anandhi,** …… and Caberlotto, L. "Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease." Alzheimer's & Dementia 12.6 (2016): 645-653.

Contents

List of Abbreviations

NDD	Neurodegenerative diseases
AD	Alzheimer's disease
CNS	Central nervous system
FTDP-17	Frontotemporal dementia with parkinsonism-17
NIH	National institute of health
MCI	Mild cognitive impairment
PD	Parkinson disease
MS	Multiple sclerosis
APOE	Apolipoprotein E
PSEN1	Presenilin E
APP	Amyloid precursor protein
PPI	Protein protein interaction network
GWAS	Genome wide association studies
GRN	Gene regulatory network
TF	Transcription factor
DAG	Directed acyclic graph
GEO	Gene expression omnibus
SGD	Saccharomyces genome database
HGP	Human genome project
MIAME	Minimum information about microarray experiment
MIAPE	Minimum information about protein experiment
URI	Unified resource identifier
XML	Extensible markup language
RDF	Resource description framework
OWL	Ontology web language

GO	Gene ontology
FMA	Foundational model of anatomy
RadLEX	Lexicon of radiological information
SNOWMED-CT	Systematized nomenclature of medicine-clinical terms
DICOM	Digital imaging and communications in medicine
EMAP	Edinburgh mouse atlas project
SBML	Systems biology markup language
URN	Uniform resource name
SPARQL	SPARQL protocol and RDF query language
LODD	Linked open drug data
ADNI	Alzheimer disease neuroimaging initiative
AIBL	Australian imaging, biomarker and lifestyle flagship study
SveDem	Swedish dementia registry
MCI	Mild cognitive impairment
EOAD	Early onset Alzheimer disease
LOAD	Late onset Alzheimer disease
PET	Positron emission tomography

GLOSSARY

Apoptosis

Programmed cell death occurring in multicellular organisms

Atrophy

A physiological process resulting in breakdown of tissues, involving apoptosis

Blood-oxygen-level dependent contrast imaging (BOLD)

A technique used in imaging technologies for observing different areas of the brain or other organs.

Connectome

A comprehensive, descriptive map of neural connections in the brain

Endocytosis

A type of transport mechanism that moves particles, such as large molecules, cellular debris or even whole cells into a cell.

Frontotemporal dementia with parkinsonism-17 (FTDP-17)

A form of major neurodegenerative disorder characterized by loss of neuronal cells in frontal and temporal lobes. It is also known as frontotemporal dementia

Neuroinflammation

An automatic, initiated response to infection, injury, toxic metabolites in the brain

Oxidative stress

An imbalanced state between free radicals and antioxidants in the body

Pick's disease

A rare type of age-related dementia affecting frontal lobe of the brain resulting in speech problems (aphasia), behavior difficulties ultimately leading to death

Progressive neural palsy

A rare brain disorder that affects movement, walking (gait), speech, swallowing, vision, mood, and behavior.

SCAIView

An information retrieval system for Life Sciences developed in Fraunhofer Institute of Scientific Computing and Algorithms (SCAI). For details see www.scaiview.com

tranSMART

An open-source knowledge management platform for clinical and translational research. For details see http://transmartfoundation.org

CHAPTER 1

Introduction

1.1 PREAMBLE

This dissertation is not a single-scale research project, but instead, it is a collection of the diverse research questions that I have explored in the past four years of my thesis.

The central objective of this dissertation is concerned with multi-scale, integrative modeling of neurodegenerative disease. Although there are many ways of approaching this objective, this work primarily focuses on three significant aspects through which it can be addressed. They are as follows:

- A state-of-the-art integrative framework for assembling qualitative and quantitative data from a broad spectrum of data resources for answering complex biological questions across all scales of the disease domain
- Essential role of biomedical ontologies for explicit characterization of a specific domain of interest and support for the formalization of heterogeneous data resources
- Multi-scale understanding of a disease mechanism using interoperable clinical/imaging ontologies and molecular data

The introduction of this thesis pertains to all of the preceding aspects.

1.2 HUMAN BRAIN: THE MOST COMPLEX AND FASCINATING BIOLOGICAL ENIGMA OF THE UNIVERSE

The human brain is the most remarkable, and correspondingly complex, organ of the human body. At about 1.4 kilograms, it accommodates one hundred billion nerve cells/neurons. For this reason, brains are compared to man-made computers for their astounding ability to not only process and communicate information, but also regulate unconscious bodily actions such as digestion and breathing [1]. According to the physicist, Roger Penrose,

"If you look at the entire physical cosmos, our brains are a tiny, tiny part of it. But they're the most perfectly organized part. Compared to the complexity of a brain, a galaxy is just an inert lump."

Although the fields of neuroscience and neuroanatomy have substantially evolved over the years, unraveling the structural and functional complexity of the brain seem to remain an unattainable goal. This is partly because (i) our knowledge of neuroanatomy remains far from complete and (ii) the brain's temporal, topological and spatial multi-scale networks give rise to elaborate molecular, cellular, and neuronal phenomena that regulate the basis of cognition ([2],[3]).

Human Brain Architecture

Over the past decade, the neuroimaging community has witnessed a paradigm shift attracting considerable attention from many disciplines of scientific investigation. Functional neuroimaging techniques provide a multi-level view of the brain's intrinsic connections between neurons and individual brain regions, collectively known as connectomes ([4],[5]). The term "*connectome*" defines the concept of representation

of the human brain in the form of a complex, large-scale neural network ([6],[7]). Structural connectomes define the length and strength of the projections between brain regions whereas functional connectomes represent the alterations of anatomical projections based on functional properties such as cognition and behavior ([8],[9], [10], [11]).

Various scales often characterize the hierarchical organization of these complex structural and functional networks. These are chiefly spatial, temporal as well as topological properties of the brain network. The spatial scale represents the granularity at which the nodes and edges are defined, ranging from individual cells/synapses (represented as nodes) to large-scale brain regions and the strength of interaction between the brain regions (edges) ([12],[13],[14]). A temporal scale represents the brain network at a given time, ranging from millisecond to the evolutionary changes occurring over several generations. The topological level of a brain network can be represented either as a single node or as a whole network, depending upon the region of interest. Altogether, the topological, spatial and temporal scales represent the axes of a three-dimensional space for any analysis performed on brain networks ([15], [16], [17]). Understanding the complexity of the human brain through connectomes is essential to elucidate the pathologies of complex brain related neurodegenerative disorders, such as Alzheimer disease (AD), thereby contributing to an arduous goal of developing enhanced diagnoses and treatment options. Until now, studies have primarily focused on single-scale architectures (the aforementioned spatial, temporal and topological scales, respectively).

However, newly emerging studies are primarily concerned with bridging these scales to one another to truly recognize the multi-scale, multi-modal nature of the human brain (Figure 1).

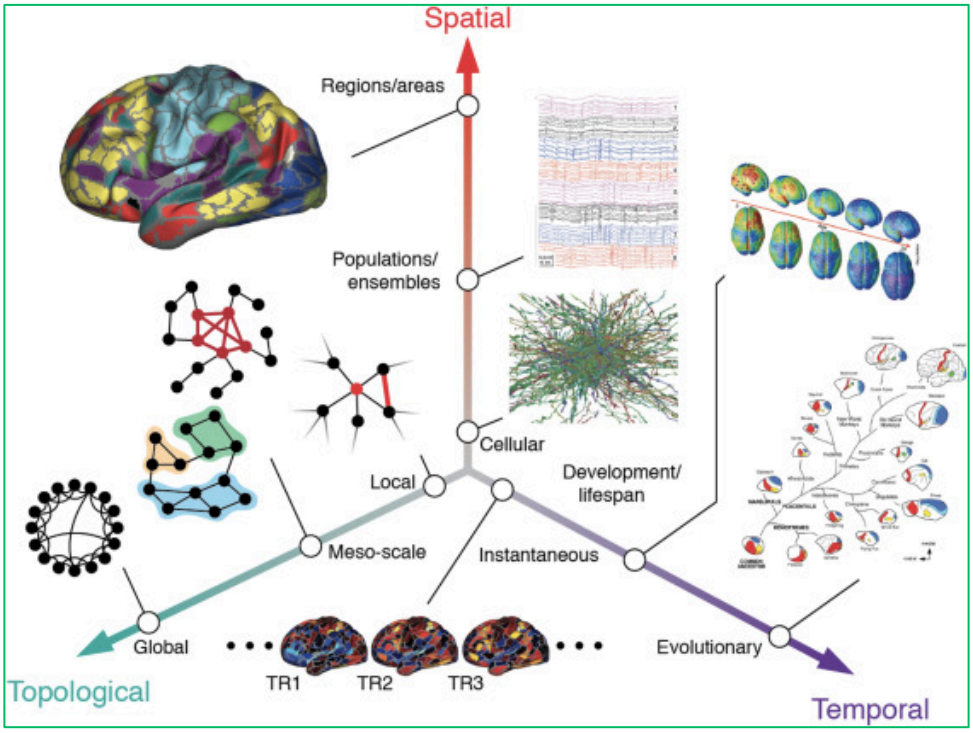


Figure 1: Three-dimensional scaling of the whole brain. The figure represents the various types of organizing brain networks across spatial, topological and temporal scales ranging from individual nodes to large-scale networks (Figure taken from [2], Last accessed on 30.01.2018)

1.3 EXPLORING THE COMPLEXITY IN NEURODEGENERATIVE DISEASES AND DEMENTIA

Neurodegenerative diseases (NDDs), commonly characterized by an unrelenting deterioration in cognitive ability and capacity for independent living, has become an alarming healthcare priority across the globe. They embody a vast spectrum of heterogeneous disorders, which ultimately result in neuronal damage and death due to many anatomical and functional complications. Amongst NDDs, Alzheimer's disease (AD) and Parkinson's disease (PD) have been identified to be increasingly hazardous concerning human suffering and economic cost ([18],[19],[20]). According to a recent collaborative study between the World Health Organization (WHO), the World Bank and the Harvard School of Public Health, by the year 2020, dementia and other NDD will be the eighth leading cause of disease burden for developed countries (([21],[22]). Taking into consideration an aging population worldwide, the WHO has estimated that NDDs will become the world's second leading cause of death overtaking cancer by the year 2050 ([23], [24]). Although this fact is acknowledged by governments worldwide leading to significant investments in research programs to combat NDD, promising solutions to cure NDD remain lacking due to the heterogeneity of these diseases ([25],[26],[27],[28]). Therefore, it is imperative that the research community and governmental organizations systematically work towards early diagnosis and preventive measures against NDD by aligning the goals of the research community with those of the society.

1.3.1 Alzheimer disease: A multifactorial challenge

Amongst the common NDDs, AD is considered the most detrimental, affecting over 5 million people worldwide ([29],[30],[31]). The etiology of AD includes loss of

neurons, the presence of abundant neurofibrillary tangles and extracellular deposits of amyloid plaques ([32],[33],[34]). Neuropathological observations conducted more recently suggest that the accumulation of amyloid beta is the initial trigger for various biochemical and pathophysiological alterations including neuroinflammation, synaptic dysfunction, astrogliosis, tauopathy and ultimately cell death ([35],[36],[37],[38]). Although it is viewed as a first order neurodegenerative process, the terminal phase of AD patients is most often associated with a significant lapse of social and behavioral interactions, which are systemic manifestations beyond those occurring in the central nervous system (CNS) ([39],[40]). While the rate of physical decline has been observed in varying degrees during the earliest stage of AD, it is noteworthy that these are visible indicators which arise before the presence of any structural, functional and metabolic alterations. Furthermore, the classical diagnosis of the existence of hyperphosphorylated tau in AD is also found to be a hallmark for other related neurodegenerative diseases such as frontotemporal dementia with parkinsonism-17 (FTDP-17), Pick disease and progressive neural palsy ([41],[42],[43]). Considering all of the above clinical indications, AD is regarded as a "multi-factorial syndrome" due to the involvement of multiple factors leading to the dysregulation of critical pathways on various levels resulting in varied clinical manifestations (Figure 2). Less than 1% of AD cases are attributed to genetic mutations of three proteins, namely amyloid precursor protein (APP), presenilin 1 (PSEN1) and presenilin 2 (PSEN2) ([44],[45],[46]). About 99% of reported AD cases are caused by sporadic forms of AD with several different etiopathogenic mechanisms, such as neuroinflammation, head trauma, and diabetes, but are not associated with any mutation ([47],[48],[49],[50]). Moreover, the presence of allele APOE4 has been found to increase the risk of the disease by several folds.

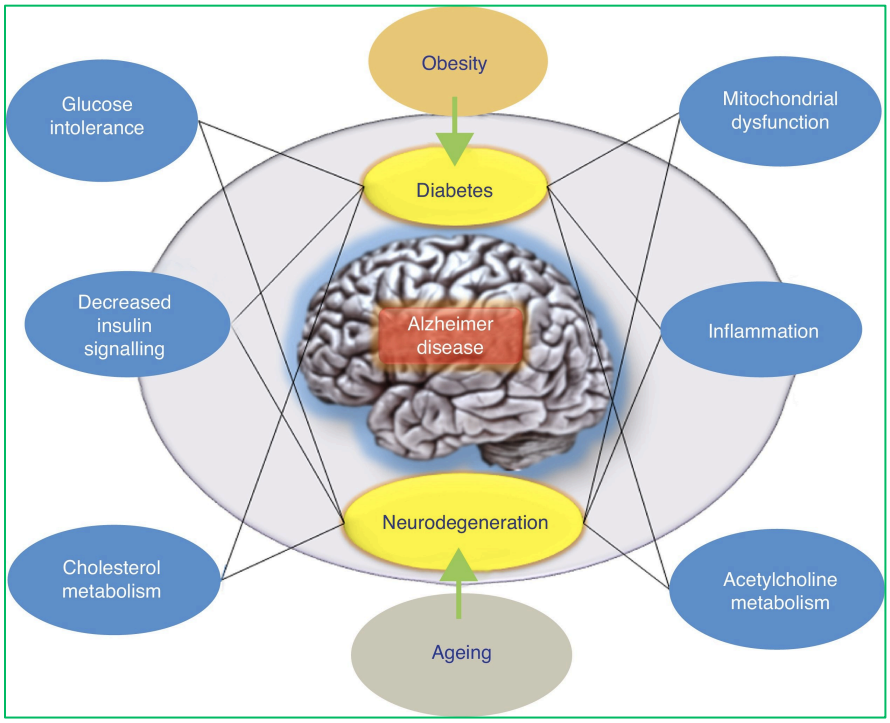


Figure 2: Multiple factors and processes linked to AD. This figure illustrates the various factors influencing AD namely genetic, metabolic and behavioral factors. (Figure is taken from [50], Last accessed 30.01.2018)

The histopathological examination of amyloid plaques and neurofibrillary tangles (NFT) of hyperphosphorylated tau are not unique to AD patients. Studies on healthy individuals have also revealed the presence of amyloid plaques in equal proportions to those seen in AD cases ([51],[52]). Similarly, the presence of NFT of hyperphosphorylated tau without amyloid plaques has been observed in FTDP-17 and Pick disease. Besides, non-demented, active individuals are also at risk of harboring the underlying pathology which could eventually lead to dementia ([53],[54]). These observations indicate that the various biological and etiological processes driving the clinical manifestations of AD can differ distinctly from patient to patient ([55],[56]).

While dealing with such multifactorial, complex neurodegenerative diseases such as AD, it is of utmost importance to obtain profound knowledge about the various aspects of etiopathogenic mechanisms ([28],[57]). Such insight can not only lead to the identification of diagnostic biomarkers for early detection of the disease but also offer numerous possibilities for developing rational therapeutic drugs ([58],[59]). The development of drugs that can provide a cure requires an accurate diagnosis of the pertinent AD sub-type that takes into account clinical evaluations such as a patient's medical history, physical examination, neuroimaging and neuropsychological assessments, and potentially, a postmortem examination for the presence of NFT or amyloid plaques ([60],[61]). These clinical indicators are collectively termed "biomarkers" which can be defined as an objective measure of a biological and pathogenic process and serve as an indicator of health or disease severity. An ideal biomarker would help in distinguishing AD from other types of dementia such as Mild Cognitive Impairment (MCI), Vascular Dementia (VaD) and Frontotemporal Lobe Dementia (FTLD), to name a few ([62],[63],[64]).

1.3.2 Diagnostic pre-clinical biomarkers for patient stratification in AD

a) ***Fluid biomarkers***

The cerebrospinal fluid (CSF) biomarkers are considered to be one of the well-established intrusive as well as invasive diagnostic biomarkers for the early detection of AD ([65],[66]). They are intrusive because the CSF is obtained through a lumbar puncture which aggrieves elderly patients with nausea, severe backache and dizziness ([67],[68]). However, CSF biomarkers are imperative in tracking disease progression because they are in direct contact with the brain and therefore, any minor physical or biochemical changes in the brain would inherently reflect in the altered composition

of the CSF ([69],[70]). Biomarkers such as amyloid-beta 42 (Aβ), total-tau (t-tau) and phosphorylated tau (p-tau) in the CSF are core indicators for characterizing patients as having early AD. It has been observed that patients with AD have reduced levels of CSF Aβ (1-42) in comparison to healthy individuals due to insufficient clearance of amyloid-beta from the brain ([71],[72]). This is hypothesized because the formation of Aβ (1-42) through the amyloidogenic pathway results in the accumulation of amyloid beta in the brain whereas activation of a non-amyloidogenic pathway in healthy individuals results in no such formation of Aβ (1-42) ([73],[74]).

Similarly, NFT and dystrophic neurites are other pathological indicators of AD. The primary constituent of NFTs is a hyperphosphorylated form of tau protein. These tau proteins are found to be synthesized within neurons, and, in the healthy state, play a role in enhancing axonal transport and providing stability to microtubules. However, during AD progression, tau proteins are hyperphosphorylated and subsequently disassociate themselves from microtubules, impairing axonal transport, and resulting in increased levels of p-tau in the CSF of AD patients in comparison to healthy controls. Apart from that, tau proteins have also been found to be a useful prognostic biomarker in tracing the progression from MCI to AD due to a consistent increase in CSF tau levels of patients during the disease ([75],[76],[77]).

The blood-based biomarkers are another type of fluid biomarkers that are easily accessible and well suited for repeated sampling, when compared to the CSF biomarkers as mentioned above ([78]). Although the blood biomarkers are considered as a gold standard for the detection of disease progression, how these blood-based biomarkers can effectively measure pathological changes in the brain remains unclear.

Blood plasma has been observed to contain a wide range of low-level protein concentrations ([79],[80]). However, the change in the level of the blood components, such as albumin and immunoglobulin, is detected in minute scales. Many studies have projected blood biomarkers, namely clusterin/apolipoprotein J, are involved in the clearance of cellular debris as an outcome of apoptosis in healthy individuals ([81]). Conversely, the increased concentration of clusterin in the blood denotes a higher accumulation of amyloid-burden in early stages of AD.

b) ***Circulatory biomarkers***

microRNAs (miRNAs) are a small group of non-coding RNA molecules which are mainly responsible for regulating the silencing of post-transcriptional genes. Approximately 1-4% of human genes are known to encode miRNAs, and each of the controls 200 mRNAs ([82],[83]). miRNAs are abundantly found to be circulating in various bio-fluids such as plasma, urine, tears, saliva, and CSF, and are thus considered to be highly reliable diagnostic biomarkers ([84],[85]). The dysregulation of miRNAs in the blood chronicles the various stages of disease progression in AD as well as other neurological disorders, which can be tracked experimentally by miRNA profiling technologies. hsa-miR-106, hsa-miR-153, hsa-miR-101, hsa-miR-29, and hsa-miR-107 have been identified to be prominent AD biomarkers that target APP and BACE1 proteins as they directly downregulate amyloid production in the brain ([82]). The increased identification of miRNAs for specific biological processes underscores the potential therapeutic role of miRNAs for AD.

d) ***Neuroimaging biomarkers***

The advent of state-of-the-art structural and functional imaging technologies offers an unprecedented prospect to directly observe brain structure and function, non-

invasively monitor the progression of a disease over time, or track disease trajectories ([96],[97]). Various imaging technologies have been developed over the years focusing on different aspects of the brain and its complexity. Similarly, several recently conducted longitudinal studies have revealed that imaging measures could potentially be used for distinguishing patient sub-groups/cohorts between those individuals who are at an elevated risk of AD through genetic mutations and those patients with MCI ([98],[99]).

Magnetic Resonance Imaging (MRI) is one of the most popular non-invasive imaging techniques that are used for observing structural alterations of the brain. Numerous studies have established that MRI techniques successfully demonstrated the decline from a normal state, to MCI and ultimately to AD through structural modifications of the medial temporal lobe resulting in atrophy, an early risk indicator of AD ([100],[101],[102]). Apart from the visualization of altered structural changes, MRI studies have also been used to assess volumetric changes of particular brain regions. These methods are very efficient in distinguishing them between those individuals with normal cognition and MCI from AD patients.

Diffusion Tensor Imaging (DTI) is a more recent and relatively advanced MR technique that helps in tracing the fiber tracts of the brain as well as the non-random movement of water molecules, known as “Brownian motion” or simply, “diffusion” ([103],[104]). These techniques allow for the detection of microstructural changes that can occur in fiber tracts, which cannot be so readily done using traditional MR techniques. These microstructural alterations prove to be an excellent indicator for predicting the progression of MCI to AD ([105],[106]).

Functional MRI (fMRI) is another type of MRI technique that is helpful in detecting functional abnormalities of the brain such as alterations in cerebral blood flow and blood oxygenation levels, which directly correspond to a cognitive shift in a patient's task performance ([107],[108],[109]). The fMRI technique can detect increased blood flow via blood oxygen level dependent contrast imaging (BOLD) during a task performance done by healthy individuals. A delayed BOLD response indicates less coordinated activity in some regions of the hippocampus of MCI and AD patients. Taken together, fMRI provides powerful tools to investigate brain activation patterns of early detection, classification and prediction of AD ([110],[111]).

Positron Emission Tomography (PET) is an advanced neuroimaging technique that allows for the quantification and measurement of physiological processes such as glucose metabolism and other neurotransmitter functionalities. The radioisotopic tracers used in PET imaging help in tracing the altering state of physiological processes such as the decline in cerebral glucose levels of a particular brain region. These tracers bind to pathological molecules, such as amyloid beta, and thereby, are detected in surplus in AD patients. These surrogate markers are known to be potentially sensitive in detecting the earliest changes that could occur in the progression of the disease ([112],[113]).

Single Photon Emission Computed Tomography (SPECT) is one of the most advanced molecular imaging techniques for uncovering even the slightest chemical changes, which occur prior to any structural alterations of the brain. The SPECT radioactive tracers are highly lipophilic which facilitates the easy penetration of the blood-brain

barrier (BBB) ([114],[115]). SPECT techniques are widely used in clinical diagnosis for differentiating between (i) frontotemporal dementia and dementia with Lewy bodies as well as (ii) MCI and AD through cerebral perfusion patterns [116].

Despite the fact that clinical and molecular biomarkers have proven to be useful for the stratification of disease sub-groups, the use of CSF and imaging biomarkers is surprisingly undervalued in clinical diagnosis when compared to its usability in clinical trials and academic research [66].

This is because minimal efforts have been invested towards a multi-modal and integrative approach to different biomarkers. Furthermore, it has been particularly challenging to link the findings of molecular biomarkers with early disease progression and the real potential of imaging biomarkers in elucidating disease progression has also yet to bear fruitful results.

1.3 Moving beyond biomarkers towards integrative mechanistic modeling

In spite of the massive advancement in novel biomarker discovery approaches, the current diagnostic yardstick for clinical AD remains to be the abnormal accumulation of amyloid peptides in the brain. However, it is still unclear whether the collection of amyloid peptides or tau protein formation influences cognition and behavior or if they are merely the byproducts of a secondary etiopathogenic effect such as neuroinflammation or oxidative stress to name a few ([117],[118]). One alternative is taking a network-based approach, which has thus far been instrumental in systematically integrating and interpreting the context between various biomarkers which are represented as the individual “points/nodes” of the network ([119],[120]).

The idea of a network-based approach unfolds from the premise that alterations cause complex diseases such as NDD in more than one biomarker or a pool of candidate biomarkers comprising multiple biological pathways ([121],[122]). Here, the network consists of "nodes," that represent individual biomolecules such as genes, proteins, miRNAs, non-coding RNAs, drugs or even diseases and "edges," which represent a wide range of interactions including physical, genetic, metabolic and co-expression, as illustrated in Figure 3.

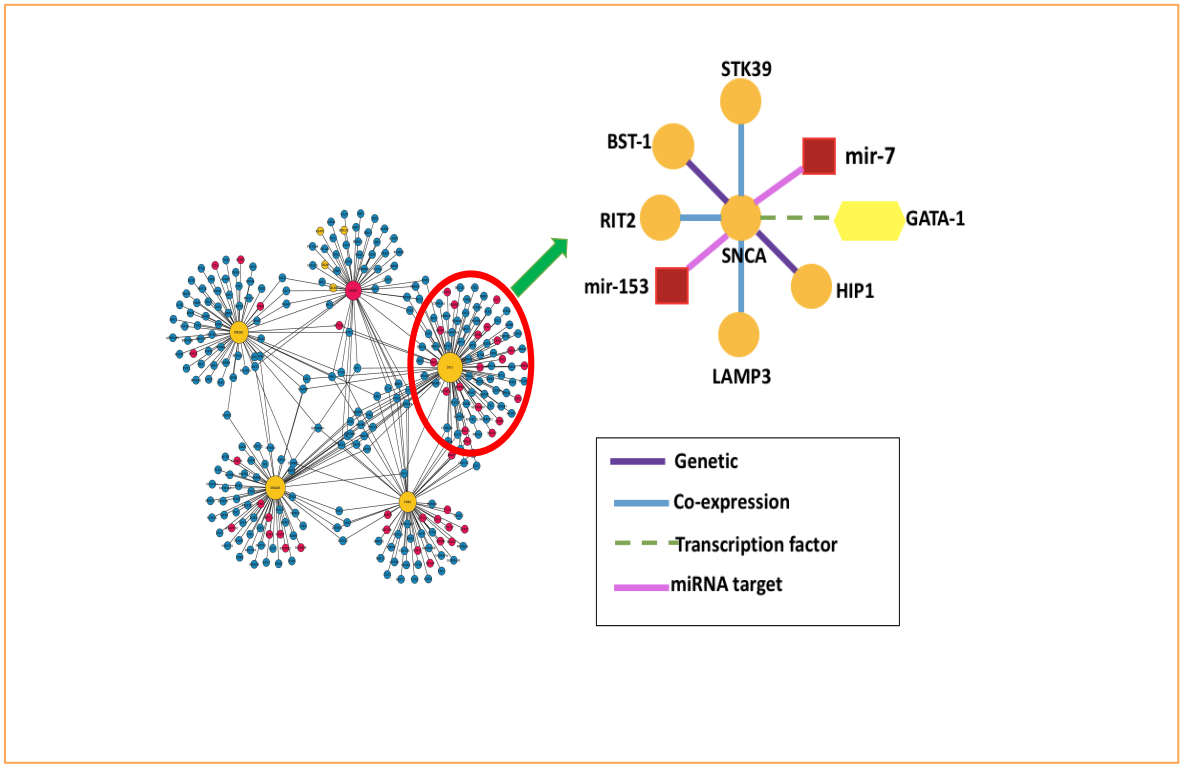


Figure 3: Representation of a biological network with multiple interactions in a disease context. The right side of the figure represents the hub of a potential gene SNCA extracted from the network. The colors of the lines represent the types of interactions between the nodes and the strength of the associations between them (Figure adapted from [123], last accessed 30.1.2018).

1.4 Pathway identification in AD through network-based approaches

The success of network-based approaches in precision medicine can mainly be accounted for by their ability to provide scientific rationales to uncover unexpected relationships between and within diseases, and they have been especially useful in deciphering molecular underpinnings in the context of AD ([124],[125]). As an example, this approach has been utilized to identify and prioritize novel candidate susceptibility genes based on the principle that genes that are associated with a disease often participate in shared biological pathways which may be highly interconnected in a network ([126],[127]).

The roles of the genes, *PDCD4* in regulating neuronal apoptosis and *ESCIT* in oxidative stress and mitochondrial dysfunction, are considered to be significant breakthroughs of this approach ([128],[129]).

Moreover, the combinatorial network studies of metabolomic and transcriptomic data have revealed various dysregulated pathways associated with AD, namely the MAPK/ERK pathway, along with the downregulation of genes such as *CYC3* and *p53* ([130],[131]). Similarly, a study by Thomas and Hallock in 2012 mechanistically deciphered that the cause of increased APP production in the brain was due to the disruption of a clathrin-mediated pathway in the context of AD ([132]).

Network level differential co-expression analysis (DCA)

Network level DCA is another advancement in network biology in which there lies the possibility of identifying causal factors of the disease through differential co-expression networks, in contrast to the traditional method of differential gene expression. The advantage of DCA over differential gene expression analysis is that it

can help in deducing the co-expression patterns of genes even in the absence of differential gene expression. The DCA approach has been successfully used in the identification of candidate regulators (*APBA2*, *SV2A,* and *FYN*) of Late Onset Alzheimer Disease (LOAD) in APP endocytosis ([133],[134]).

Weighted gene co-expression network analysis (WGCNA)

WGCNA is a statistical network approach that is useful in detecting highly interconnected modules from large networks. The principle of WCGNA is that the modules of highly co-expressed genes participate in common biological processes. WCGNA based studies have proven instrumental in NDD research. This is especially true in uncovering modules related to microglial signaling cascades in LOAD, in which there is a significant up-regulation of the gene *TYROBP*, a potential biomarker for enhancing microglial phagocytosis in clearing amyloid beta from the brain ([135],[136]).

Genome-wide Association Studies (GWAS)

GWAS have gained considerable momentum over the last decade in elucidating statistically robust associations between genetic variants and clinical phenotypes in an unbiased fashion. The advantage of GWAS is that they have conceded many significant findings that have been validated by independent studies with consistent accuracy in complex diseases including AD [137]. Naz et al. (2017), in their recent publication, have demonstrated the use of a GWAS approach to reveal molecular mechanisms with "genetic traits" based on the shared etiology of two diseases. The integrative approach established in work has resulted in the identification of shared pathophysiology between AD and PD [138].

Network biology and the approaches that subsequently followed offer an excellent platform to integrate and investigate the molecular underpinnings of complex diseases comprehensively. These approaches complement high-throughput gene profiling studies in bridging the gap between traditional and translational research, the combination of which can facilitate the accuracy of personalized medicine.

1.5 Role of ontologies and terminologies in semantic integration

Ontologies and terminologies can be defined as single taxonomic and knowledge representation schemas and are considered to be optimal solutions for biological interoperability [139]. They are conceptual models that aim to support consistent and unambiguous knowledge sharing and provide a formalized framework for knowledge integration. The field of text mining has predominantly exploited the use of ontologies as a means to retrieve domain-specific knowledge from literature and has further gained popularity in biomedical research. For instance, Gene Ontology[1] (GO) was one of the primordial ontologies developed in the field of biomedicine for producing a dynamic, controlled vocabulary for representing knowledge of all genes and proteins in Eukaryotes. Similarly, several ontologies for varying necessities, ranging from cell cycle ontology to a pathway and event ontology, have emerged and become an integral part of biological research [140].

Disease Ontologies such as SNOMED-CT[2], International Classification of Diseases (ICD[3]) and the human disease ontology contain human disease concepts for supporting the healthcare community to enrich the knowledge behind diseases.

[1] http://www.geneontology.org/

[2] https://www.snomed.org/snomed-ct

[3] http://www.who.int/classifications/icd/en/

However, these ontologies provide a high-level concept representation and do not provide the required granularity and depth of domain-specific concepts. This holds true for all diseases, especially NDD. ([78],[141]). Malhotra et al. (2014) have addressed this with the development of the Alzheimer Disease Ontology (ADO) to cover various aspects of AD, which includes concepts such as clinical, etiological and underlying cellular and molecular mechanisms [142]. Similarly, the Parkinson Disease Ontology (PDON) was developed for representing knowledge surrounding PD with concepts ranging from molecular biology to clinical readouts [143].

By cause of efficient knowledge retrieval using ontologies, biomedical databases are increasingly becoming semantically equipped for dynamic data integration and interoperability.

1.5.1 Tackling the challenge of data integration in systems biology

The advent of network biology has given rise to the unprecedented accumulation of high-throughput data from sources including genomics, proteomics, and metabolomics, amongst other. These data resources have not only provided a platform to seek answers for the most fundamental biological questions but also facilitate the formulation of novel hypotheses that arise from the accumulated wealth of data. A vast number of genes, proteins, enzymes, pathways and other biomolecules are identified, sequenced and stored in public repositories known as databases ([144],[145]). These public databases served as the initial integrative framework that provided the platform to store, organize and integrate the vast amounts of data generated through reliable experiments. Such databases ensure the reproducibility of scientific experiments and analysis and facilitate better interpretation of scientific

discoveries. On this note, the biological databases can be broadly classified into (i) sequence, (ii) structure and (iii) pathway databases ([146],[147]).

Sequence databases are exclusively dedicated to capturing nucleotide sequences and their associated biological and bibliographic information. The major public databanks for sequence-related information are GenBank[4], European Molecular Biology Laboratory (EMBL[5]) and DNA Databank of Japan (DDBJ[6]) ([148],[149]).

Ensembl[7] is another primary database that serves to annotate the human genome and predict new genes automatically. Databases such as ArrayExpress[8] and Gene Expression Omnibus (GEO[9]) are officially affiliated with the microarray gene expression data society that ensures the quality of data submitted to the repository ([150],[151]).

Apart from human genome sequences, there are other databases, which are dedicated to capturing sequence-based information from other organisms such as mouse in Mouse Genome Database (MGD[10]), and yeast in Saccharomyces Genome Database (SGD[11]) for related gene expression data ([152], [153]).

---

[4] https://www.ncbi.nlm.nih.gov/genbank/

[5] https://www.embl.de/

[6] https://www.ddbj.nig.ac.jp/index-e.html

[7] https://www.ensembl.org/index.html

[8] https://www.ebi.ac.uk/arrayexpress/

[9] https://www.ncbi.nlm.nih.gov/geo/

[10] http://www.informatics.jax.org/

[11] https://www.yeastgenome.org/

Structural databases comprise of those databases that are exclusively dedicated to capturing protein related data to understand the role of protein functions in complex mechanistic processes.

Protein databases such as Protein Data Bank (PDB[12]), Pfam[13] and UniProt[14] are dedicated to the collection and storage of protein structures ([154],[155],[156]). The CATH[15] database provides information on the hierarchical classification of protein structures. Proteins do not act independently but rather as a network of complicated molecular interactions [157]. Experimental techniques such as Yeast two-hybrid (Y2H) help in capturing the physical interaction of proteins and such information is primarily stored in databases such as BioGRID [16], MIPS [17], and STRING[18]([158],[159],[160]).

Pathway databases are essential for successfully quantifying biological entities. Kyoto Encyclopedia of Genes and Genomes (KEGG[19]) has been the primary data

---

[12] https://www.rcsb.org/

[13] https://pfam.xfam.org/

[14] http://www.uniprot.org/

[15] http://www.cathdb.info/

[16] https://thebiogrid.org/

[17] http://mips.helmholtz-muenchen.de/proj/ppi/

[18] https://string-db.org/

[19] http://www.genome.jp/kegg/

resource for storing computerized knowledge on molecular interaction networks, predominantly comprising of metabolic pathways [161].

Ingenuity pathways knowledge database consists of data about gene regulatory and signaling pathways for various organisms including human, mouse, and rat [162]. Reactome[20] is a far more advanced and dynamically designed database that captures cell signaling and metabolic pathways. Their framework facilitates the ability to superimpose quantitative data that can be highly useful for experimental validation [163].

Early initiatives, such as the Human Genome Project (HGP) which generated over 2 million data records per day, signified the importance of using databases not just as a storehouse but as a systematic framework for interlinking inter-related data resources [164]. Integration of data resources was perceived to have multiple advantages such as:

- Providing a collective view of various heterogeneous resources in one location
- Modeling biological systems at cellular, molecular, metabolic and systemic levels
- Assimilation of literature and data-driven knowledge
- Querying of the knowledge resource to answer fundamental questions
- Formulation and test of novel hypotheses about disease mechanisms [165]

1.5.2 Large-scale initiatives towards biological data warehousing

[20] https://reactome.org/

Integration of disparate sources containing heterogeneous biological information has been a challenging task in the field of bioinformatics, though the outcomes of these approaches have contributed to many scientific discoveries over recent years. Early public data repositories would often focus on providing only a single type of data, for example, the Bimolecular Interaction Network Database (BIND) for retrieving molecular interactions, and BRENDA [21] for retrieving enzyme related data ([166],[167]). However, current databases are not entirely suitable for sophisticated, system-wide research, which requires the integration of a large number of distinct existing resources. The bioinformatics field has seen a significant paradigm shift from single source databases to complex warehousing systems which aid in accommodating the wealth of "omics" data [166]. BioMart[22], BioXRT[23], and InterMine[24] are generic genomic data warehousing systems with 28 data sources exclusively designed for integrating and querying genomic data [168]. In the systems biology front, PathwayTools[25] is a rapidly growing pathway warehouse system leveraging on average 1,800 manually curated pathways from MetaCyC[26] database [169]. The most sophisticated data integration platform that was widely popular amongst the research community was DBGET[27]. It was considered to be one of the largest integrated data retrieval systems containing an array of molecular biology

[21] https://www.brenda-enzymes.org/

[22] http://www.biomart.org/

[23] http://projects.tcag.ca/bioxrt/

[24] http://intermine.org/

[25] http://bioinformatics.ai.sri.com/ptools/

[26] https://metacyc.org/

[27] http://www.genome.jp/dbget/

databases [170]. The visual representation of DBGET is in the form of an extensive web graph in which individual databases within the integrated platform are represented as "nodes" and are cross-referenced among other databases as edges, as illustrated in Figure 4.

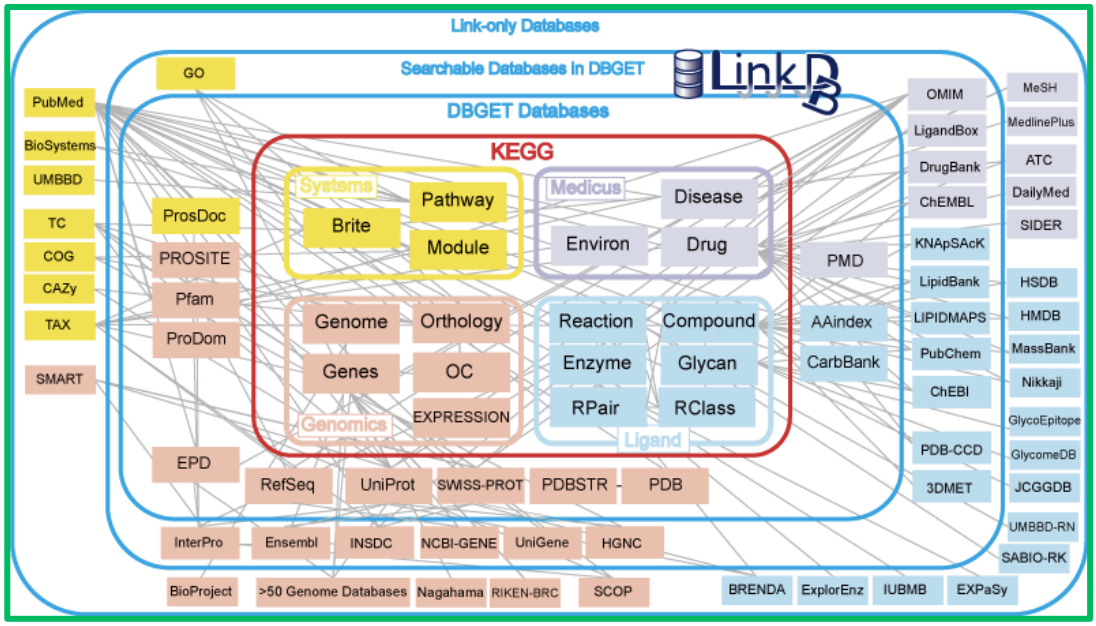


Figure 4: DBGET architecture[28]. This figure is an integrated database retrieval system for a major biological database (Link last accessed on 30.01.2018)

Numerous large-scale initiatives towards a disease-specific warehousing system have been on the rise in recent years. The cancer Biomedical Informatics Grid (caBIG) is an open source, open access information network for secure data exchange on cancer research. The warehouse system offers the necessary accessibility for collecting, analyzing, integrating and disseminating information associated with cancer research [171].

[28] http://www.genome.jp/linkdb/

The Systematic Platform for Identifying Mutated Protein (SysPIMP [29]) is a standardized data warehouse designed for studying human diseases caused by mutation. This database is an integrated resource of several known databases such as Protein Mutant Database (PMD[30]), Online Mendelian Inheritance in Man (OMIM[31]) and Human Gene Mutation Database (HGMD[32]), to name a few [172].

Concerning NDD, the Global Alzheimer's Association Interactive Network (GAAIN[33]) has provided a gateway to access the vast knowledge collection of AD research data facilitated by sophisticated analytical tools and databases. This warehouse system offers a federated network for data sharing with the community interested in dementia research [173].

Despite the several advantages of data warehousing systems, there are major hurdles with this approach as it requires continuous and often human guided management to remain updated, resulting in high-cost maintenance as well as a distortion in the quality of data over time. The factors affecting the quality of data are majorly due to syntactic and semantic anomalies. Syntactic anomalies include the challenges of incomplete data, inaccuracy, lexical errors as well as formatting issues. Semantic anomalies include data ambiguity, data discrepancy, and redundancy ([139], [174]).

[29] https://omictools.com/systematic-platform-for-identifying-mutated-protein-tool

[30] http://pmd.ddbj.nig.ac.jp/~pmd/pmd.html

[31] https://www.omim.org/

[32] http://www.hgmd.cf.ac.uk/ac/index.php

[33] http://www.gaain.org/

Lack of semantic standardization in public data repositories makes it extremely difficult to interpret the real context and meaning of data that has been derived from multiple sources or synthetically generated by software technology. Thus, without a formalized structure, the burden of knowledge interpretation as well as determining the validity and meaning of the data remains a substantial constraint among data users.

Semantic integration is a highly innovative concept that focuses on the use of metadata to describe the meaning of data, which is facilitated by standardized ontologies that enable easy interoperability of concepts from one domain to another.

1.6 Formalization of biological networks in biomedicine: an insight into semantic web technology

Current research in the life science necessitates the understanding of data at multiple levels, ranging from cells to whole biological systems, with accessibility across numerous species and various experimental conditions. To explore increasingly complex relationships, there is a constant need for leveraging and linking different types of information, while at the same time, challenges exist with fragmented resources and technologies.

The notion of the semantic web is gaining significant momentum in the life sciences as it provides the necessary infrastructure for semantic and syntactic interoperability for efficient data integration through the World Wide Web.

Sir Tim Berners Lee, who coined the term semantic web, defined it as the following:

"an extension of the current web in which the information is given well-defined meaning, better enabling computers and humans to work in cooperation."

Semantic web technologies (SWT) provide a standardized platform for expressing relationships between web resources, thereby easing the way for data and machine interoperability ([175],[176]). The concept of the semantic web can be well explained by a semantic-web architecture, commonly known as a "semantic web layer cake," as displayed in Figure 5.

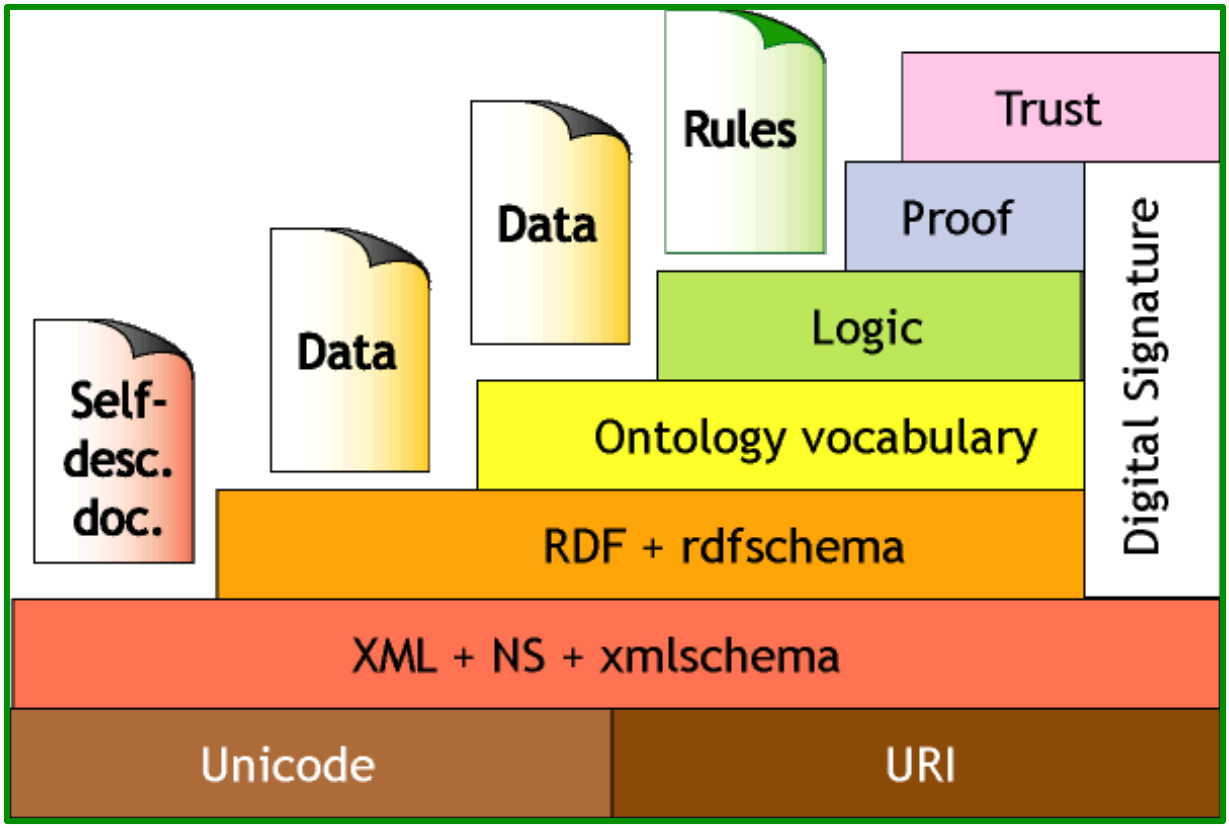


Figure 5: Semantic Web Architecture[34]. This figure enables the standards and technologies needed to build a robust semantic framework for organizing knowledge.

SWT works on three main aspects of data integration using:

[34] https://www.researchgate.net/figure/The-Original-Semantic-Web-Architecture_fig4_266375123

- Ontologies for the representation of domain-specific concepts
- Resource Description Framework (RDF) as the standardized representation of language
- Accessibility of data through the web language, SPARQL Protocol and RDF Query Language (SPARQL) ([177],[178]).

The usability of ontologies for semantic interoperability has been previously explained in detail (cf section 1.3).

RDF is the standardized language for data representation and interchange on the World Wide Web. They provide a framework for creating statements about resources and their corresponding attributes. RDF makes use of a Universal Resource Identifier (URI) to identify every element represented in the RDF model ([179],[180]). The basic structure of RDF is in the form of a statement containing three elements, namely a "subject," "predicate" and "object" collectively known as "triples" [181]. A collection of triples forms a network of interconnected nodes and edges that describes the nodes and their relationships with other nodes. A simple representation of a triple can be seen in Figure 5. The statement "*APP is encoded by an amyloid beta peptide protein*" is represented in the form of subject, predicate, and object. The literal represents the actual concept without the URI.

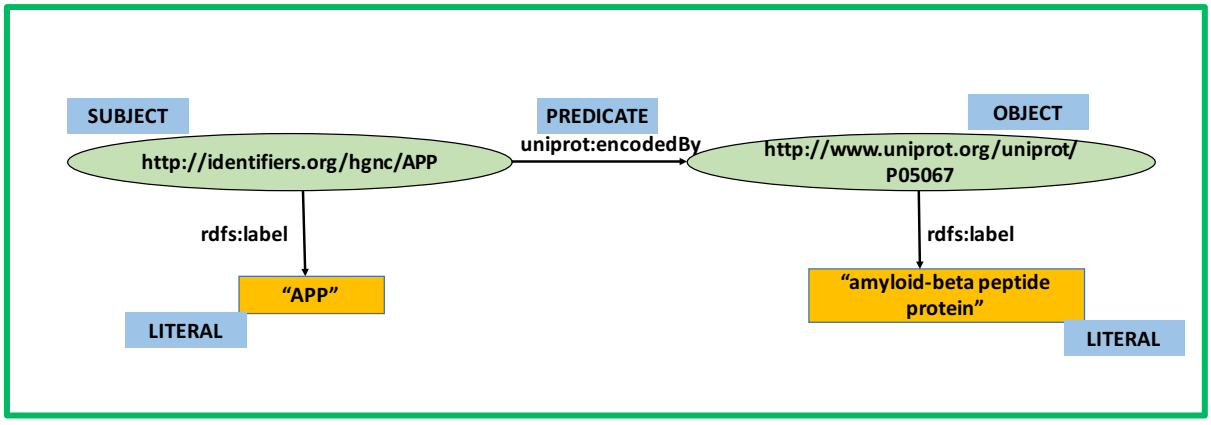


Figure 5: **Representation of a triple in RDF data model.** *This figure illustrates a simple biological statement converted into a triple format.*

The RDF Schema (RDFS) is a standardized language for the representation of data properties and is based on RDF. This is achieved with the Web Ontology Language (OWL) for formally defining semantics through the concept of description logic for semantic consistency.

Since the representation of RDF is in the form of a graph, accessing specific information is only possible through querying the graph using a specific language called SPARQL.

1.6.1 Use of the semantic web in the biomedical domain

The quickest method to search for information on websites is through efficient search engines, such as Google. The search leads to a set of HTML pages that are devoid of any context or semantics and demand human effort. An analogous scenario is present in the biomedical domain, specifically in the field of molecular medicine. The *Entrez* search engine, a core part of the NCBI portal, is one of the most extensive knowledge repositories, consisting of various databases ([78],[182],[183]). The major drawback of these search engines is that the content retrieved is not machine accessible. SWT has overcome these challenges through various methods, including upheaving the exchange of data without any loss of information, formalizing the data into computable knowledge, transforming integrated data into "smart data" for biomarker discovery and deriving novel hypotheses ([184],[185]).

BioGateway[35] was amongst the earliest RDF based platforms built to aggregate biological ontologies and other knowledge resources such as OBO Foundry candidate ontologies, GO annotation files, SWISSPROT[36] protein sets, and NCBI taxonomy. This framework acts as a single point of entrance to query the integrated knowledge using the SPARQL protocol [186].

Similarly, RDFscape is another semantic framework that was developed to facilitate biological analyses with an ontology, where reasoning power can be applied [182].

Semantic Web Applications and Tools for Life Sciences (SWAT4LS[37]) is yet another promising initiative towards creating open source, linked data. This platform brings together researchers from various sectors, namely eHealth, medical and clinical informatics, bioinformatics, chemoinformatics and systems biology to exchange ideas in the application of semantic web in health care [187].

The Bio2RDF project revolutionized life science research by demonstrating the possibility of querying the life science knowledgebase through linking large-scale publicly accessible databases through SPARQL endpoints. Bio2RDF uses RDF documents as well as a rule-based approach for creating URIs that automatically create linked data. This framework provides access to the semantically normalized RDF documents that are generated from public resources. It contains over 163 million documents from 20 data resources. Figure 6 represents the Bio2RDF framework ([188],[189]).

[35] http://www.thebiogateway.org/

[36] https://web.expasy.org/docs/swiss-prot_guideline.html

[37] http://www.swat4ls.org/

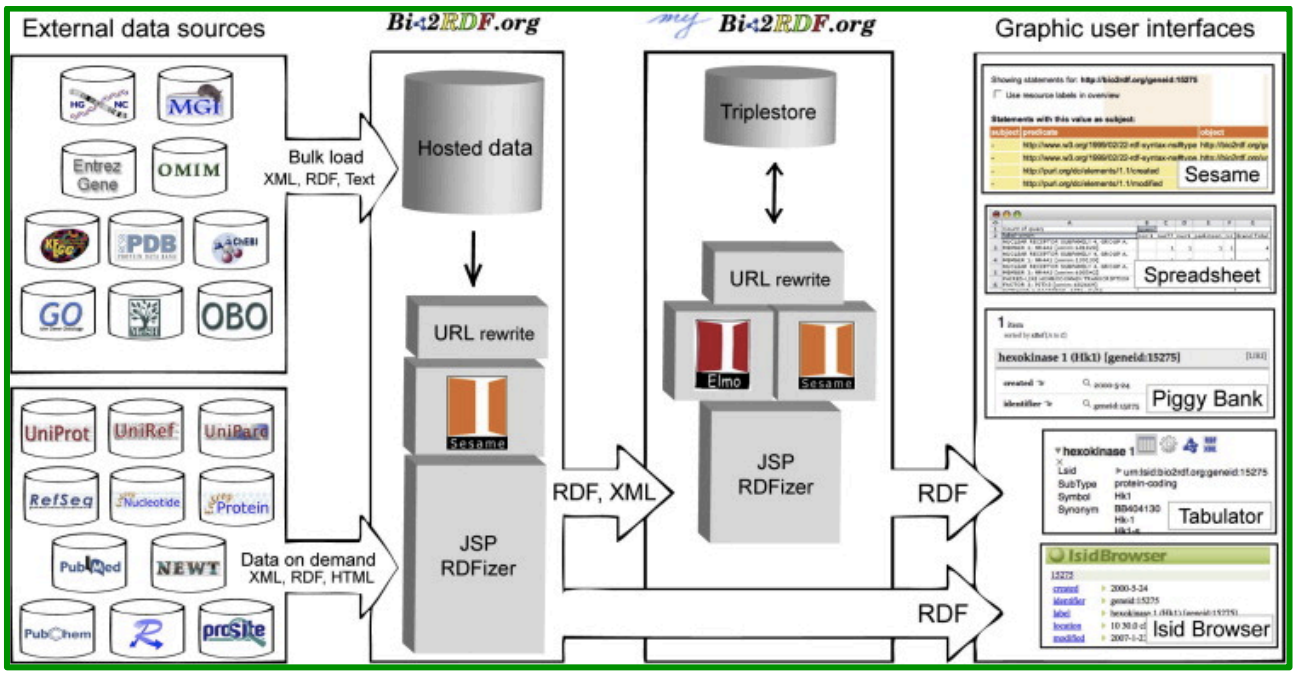


Figure 6: Bio2RDF knowledge system architecture. A schematic representation of the Bio2RDF framework, depicting the processing of external data resources to standardized RDF documents (Figure adapted from [188], last accessed 30.1.2018).

1.6.2 Use of semantic web in neuroscience

The key to unraveling novel hypotheses from integrated data resources has been well tailored for clinical research. Diseases and adverse events are evolving phenomena, and therefore, the needs for innovative strategies to understand, diagnose and treat them also continue to grow simultaneously. AlzPharm[38] was one such initiative launched to support research around NDD, especially AD actively. It is an e-neuroscience framework that was designed to answer complex neuroscience-related questions through an integrated system of databases ([190],[191]).

[38]https://omictools.com/alzpharm-tool

The Semantic Web Application in Neuroscience (SWAN) framework was built primarily using Alzforum[39] for aiding scientists in organizing, managing, sharing and comparing knowledge related to AD. SWAN has since evolved into a valuable source for hypotheses generation and experimental design ([107],[192]).

With the advent of powerful platforms coupled with the availability of experimental data, multi-scale modeling has expanded in ways that facilitate the comprehensive investigation of biological phenomena systematically. With the standardization of concepts and knowledge resources using ontologies, it is now a simple task to integrate data resources on various scales and have a biologically consistent interface between them ([193],[194]).

To fully exploit the power of semantic web or any integrative multi-scale modeling approach, it is important to first perform a check for the completeness of the available data resources ([195],[196]). One of the most significant challenges in the healthcare domain is the lack of accessibility to individual clinical records owing to the legal and ethical principles governed by hospitals. The importance of sharing and reusing data in biomedical research is invaluable as it empowers easy accessibility and interoperability of data between independent platforms in clinical practice ([197],[198]). With the prevalence of electronic health records, it is now possible to explore individual patient health records, as well as there is a substantial increase in the volume of available data for conducting further experiments. This is highly valuable concerning diseases such as NDD, where there is always a shortage of available data ([199],[200]).

[39] https://www.alzforum.org/

1.7 Emerging trend towards a transparent infrastructure for data sharing through public-private partnerships

The era of personalized medicine aspires at establishing strong links between bimolecular characterizations, progressive indications of disease and effectiveness of treatment and drug effects as a collective diagnostic solution for individual patients ([107], [201]). It is also a fact that major breakthroughs in the field of biomedicine, especially systems biology, have only transpired because of the substantial improvement in bimolecular knowledge, as well as technologies in the recent years. Research enthusiasts have welcomed the trend towards collaborative initiatives between pharmaceutical companies and research consortia as such an approach yields high research productivity, reduced research risk and fosters increased scientific innovation ([202],[203]).

In this direction, *tranSMART*[40] was one of the earliest public-private partnerships to render analytical tools invaluable for clinical and translational research. tranSMART heavily campaigns for data sharing and enables users to compare patterns of gene expression between healthy and diseased individuals, access clinical data as well as provide the possibility to integrate external public resources ([204],[205]).

The Encyclopedia of DNA Elements (ENCODE[41]) consortia is an international collaboration between a community of academics and researchers and The National

[40] http://transmartfoundation.org/

[41] https://www.encodeproject.org/

Human Genome Research Institute (NHGRI[42]) aimed towards understanding the functional elements of the human genome ([206],[207]).

In healthcare, the value of collaborative studies is beginning to impact the field of neuroscience. Indeed, there are numerous consortia in the AD research realm which aim at addressing different critical requirements and knowledge gaps through a collaborative framework.

The Critical Path (*c-Path*[43]) has been one of the pioneering initiatives of the Food and Drug Administration (FDA) which launched focus on developing new technologies and methods to accelerate drug development for AD and PD. Their earliest achievements have been towards establishing Clinical Data Interchange Standards Consortium (CDISC[44]) standards of concepts across many leading diseases such as AD, PD, Schizophrenia, Multiple Sclerosis (MS) to name a few ([208],[209]).

The current advancements in AD research owe a great deal of debt to the Alzheimer Disease Neuroimaging Initiative (ADNI[45]), one of the largest multi-site, longitudinal studies. This has created a profound impact in clinical research and holds great promise for future translational research as ADNI provides access to a broad spectrum of patient-centric records, namely clinical, genetic and imaging, as well as neuropsychological tests and bio-specimen markers through the process of healthy

[42] https://www.genome.gov/

[43] https://c-path.org/

[44] https://www.cdisc.org/

[45] http://adni.loni.usc.edu/

aging to MCI and eventually dementia. The rationale governing ADNI is that by employing an array of distinct biomarkers ranging from genetic to clinical, the best combination of biomarkers can be identified for determining the associated processes of the brain with AD ([210],[211]).

The impact that the ADNI initiative has created in the field of AD research, both at the molecular as well as the clinical level, remains indisputable. Historically, AD was considered to be caused by the accumulation of abnormally folded proteins by pathologists and molecular biologists and regarded as a clinical disease by psychologists and neurologists. Therefore, bridging the gap between these two perspectives had been an elusive goal until the advent of ADNI. The current ADNI study comprises the largest patient cohort with 400 subjects with MCI, 200 with early AD and 200 healthy controls ([212],[213]).

The Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing (AIBL[46]) is one of the largest cohort-based studies launched in Australia, shadowing the success of ADNI. AIBL aims to assist the development of essential and robust techniques through the use of biomarkers, cognitive characteristics and health and lifestyle factors [214].

All of these research initiatives were launched to formalize the need for clean, interoperable and reliable datasets to form a clear path towards early disease diagnosis and treatment. They also mark an apparent paradigm shift that the scientific community is making towards identifying reliable diagnostic biomarkers and better therapeutic intervention.

[46] http://www.aibl.csiro.au/

1.8 Emergence of clinical and imaging biomarkers: closing the gap and widening the scope of translational medicine and biomarker discovery

The therapeutic patient selection for AD clinical trials is traditionally based on the presence of either MCI or AD. In spite of high-impact studies such as ADNI, there are still no clear distinctions between the testing of therapeutic agents targeting Early Onset Alzheimer Disease (EOAD) versus Late Onset Alzheimer Disease (LOAD). Although EOAD and LOAD are the earlier and the late manifestations of AD respectively, the current biomarkers are thus far inadequate to even minimally distinguish two diseases with entirely different etiologies. Additionally, many established studies have stated that the clinical onset of AD commences 10-20 years before the actual onset of the disease. It is also to be noted that current treatment therapies only provide symptomatic relief for patients but do not halt disease progression. This lends further credence to the fact that although there are evident advancements in the generation of clinical data, there is a substantial lack of insight about the course of the disease ([215],[216],[217],[218]).

Although substantial financial and technological investments in collecting longitudinal biomarkers such as imaging and clinical readouts are necessary, they are only sufficient enough to perform retrospective data-driven associations with clinical phenotypes which, to an extent, may result in deriving empirical hypotheses ([219],[220]). However, what is essentially needed are robust mechanism-based modeling approaches where existing knowledge on multi-scales can be formally integrated into systematic infrastructures like RDF upon which corresponding multi-

level data-driven analytics can be applied for efficient and optimized drug development.

The importance of clinical biomarkers, especially neuroimaging biomarkers in the field of systems biology, can be very efficient as they play a pivotal role in linking clinical outcomes with molecular underpinnings during AD progression. The result of such clinically relevant observations can not only corroborate the molecular and genetic consequences of the disease but also provide insights into social and behavioral alterations exhibited by individual patients ([99], [221],[222]).

Thus, there is a clear need for integrative approaches which enable multi-scale modeling of both clinical and biological data with the aim of bridging the translational gap and expanding the capacity to find optimized treatment for patients.

Goal of the thesis

Multi-scale modeling of brain disorders, especially AD, provides the possibility of bridging scales of clinical importance, which is inclusive of measures spanning from molecular alterations to cognitive deformities. However, constructing such a multi-scale modeling approach for AD has yet to be achieved due to:

(i) The absence of an integrative platform to organize knowledge across all relevant scales

(ii) Lack of comprehensive, quantitative datasets of a disease

(iii) Lack of an infrastructure to normalize the available public data resources

(iv) Non-availability of longitudinal clinical data

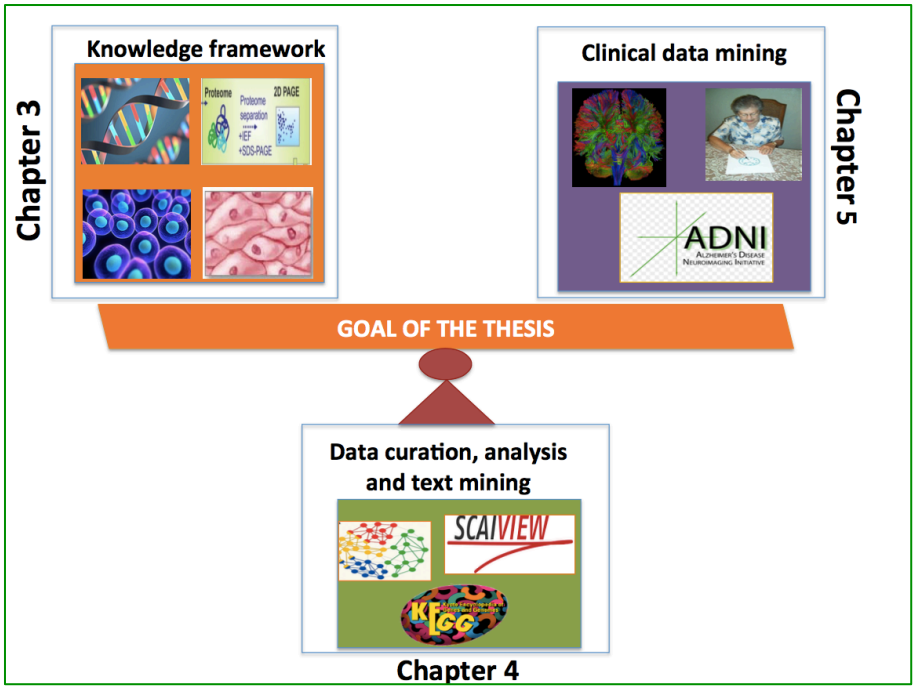


Figure 8: General workflow of this thesis

The objective of this thesis:

- ***To develop a formalized data integration platform for traversing across scales to retrieve biologically relevant mechanisms for AD***

Over the past few years, the use of biological databases by clinicians and research enthusiasts have increased manifold. This is mainly due to the emergence of web-accessible databases across all biological scales ranging from genes to brain connectivity. However, accessing and integrating essential information can be challenging due to the availability of data in varying formats and structures, the possibility of data being distributed across various public repositories which may be inaccessible, as well as the potential for poor reliability of available resources. To address these challenges, a semantic framework was designed during this work that allows for the integration of public data repositories of AD. With this semi-automated framework, it is now possible to traverse across varying data layers and query the resource to answer informative biological questions pertinent to AD.

- ***To develop a pathway inventory for multi-scale modeling underlying neurodegenerative diseases***

It has been firmly established that complex diseases are most likely caused by the alteration and dysregulation of different genes in a pathway resulting in varying manifestations of the disease. Therefore, pathways and networks provide an ideal basis for a framework to identify novel disease mechanisms and targets for drug therapy and treatment. These pathways and networks are often stored in databases such as Reactome and KEGG. Conversely, the overwhelming influx of knowledge in the form of publications and other knowledge resources have made it difficult for mining essential information from the literature. The usage of ontologies and

terminologies in biomedical research triggered substantial interest in the research community for extracting domain-specific knowledge from the vast array of literature resources. Mining context specific pathway information provides new possibilities towards the identification of novel targets. However, extraction of pathway information from the literature via text mining remains a considerable challenge as the representation of pathways in publications has been through cartoons and static representations, which are often time-consuming and non-interoperable. To address this challenge, a pathway terminology system was developed to ensure broad coverage of pathways of the neurodegenerative diseases domain.

- ***To revolutionize the mechanistic interpretation of AD through clinical biomarkers***

The recent clinical practice has seen an abundance of clinical biomarkers for early treatment and diagnosis. Clinical biomarkers, especially imaging biomarkers have gained significant recognition for predictive, diagnostic and prognostic treatment in dementia. However, these imaging biomarkers and other clinical biomarkers have so far been confined to monitor patients in clinics as well as in radiological assessments but have not yet made a significant impact in academic research. The proper inclusion of imaging biomarkers in a multi-modal approach might maximize the potential of these biomarkers to enhance our understanding of the disease and to help find effective treatments ([117],[226],[227]).

During this work, a neuroimaging terminology was developed to organize measured parameters of neurodegeneration. This terminology system helps in retrieving information relevant to neuroimaging from the literature and has been further incorporated into a computational model for understanding the relationship

between molecular processes and clinical outcomes. The overall workflow of this thesis can be seen in Figure 8.

CHAPTER 1

Heterogeneous data integration into a formalized framework and mining large-scale molecular data

INTRODUCTION

Integrative modeling approaches provide the capability to transcend boundaries between domains of varying granularity. They are highly optimized for investigating complex biological phenomena traversing multiple data domains as well as deriving novel hypotheses about complex biological processes, especially in the context of disease pathomechanisms. Constructing such a well-scalable framework requires robustness, homogenization of data across all scales as well as secures interoperability between disparate resources. However, with the indispensable growth of data resources combined with the varying quality and format of stored data, building such a heterogeneous framework is nearly impossible.

The publication presented in this chapter emphasizes on a semantic integration approach using an RDF framework for formal representation and integration of disparate data of AD. This paper particularly highlights the importance of integrating highly curated knowledge resources along with experimentally reliable data for deriving novel hypothesis about AD.

CONTRIBUTION TO THE MANUSCRIPT

Of the first four author publications presented in this thesis, the NeuroRDF paper is an exception for "shared first-author" publication in which I have equally contributed. The review and decision-making process by the Journal of Biomedical Semantics took a considerable amount of time, and we therefore agreed upon the current author list. I have made most of the contributions for the other three first author publications.

My contributions to the following paper are:

- I was entirely responsible for designing the NeuroRDF framework and generation of RDF models.

- I have performed all the experiments and data validation of this publication as well as contributed to the biological interpretation of the prioritized AD candidates.

- I have generated all the figures and tables shown in the manuscript.

- I have also taken the lead to write the following sections of the manuscript namely – Motivation, Generation of RDF models, Construction, validation and storage of RDF models, Data mining and analysis, Results and Discussion – MIF's role in AD and finally Conclusion.

- Shweta Bagewadi has contributed to the writing of the following sections of the manuscript- Abstract, Introduction, Extracting AD-specific interactions from literature, Prioritization of AD candidates and Data curation.

Iyappan et al. Journal of Biomedical Semantics (2016) 7:45
DOI 10.1186/s13326-016-0079-8

Journal of Biomedical Semantics

RESEARCH Open Access

NeuroRDF: semantic integration of highly curated data to prioritize biomarker candidates in Alzheimer's disease

Anandhi Iyappan[1,2†], Shweta Bagewadi Kawalia[1,2*†], Tamara Raschka[1,3], Martin Hofmann-Apitius[1,2] and Philipp Senger[1]

Abstract

Background: Neurodegenerative diseases are incurable and debilitating indications with huge social and economic impact, where much is still to be learnt about the underlying molecular events. Mechanistic disease models could offer a knowledge framework to help decipher the complex interactions that occur at molecular and cellular levels. This motivates the need for the development of an approach integrating highly curated and heterogeneous data into a disease model of different regulatory data layers. Although several disease models exist, they often do not consider the quality of underlying data. Moreover, even with the current advancements in semantic web technology, we still do not have cure for complex diseases like Alzheimer's disease. One of the key reasons accountable for this could be the increasing gap between generated data and the derived knowledge.

Results: In this paper, we describe an approach, called as *NeuroRDF*, to develop an integrative framework for modeling curated knowledge in the area of complex neurodegenerative diseases. The core of this strategy lies in the usage of well curated and context specific data for integration into one single semantic web-based framework, RDF. This increases the probability of the derived knowledge to be novel and reliable in a specific disease context. This infrastructure integrates highly curated data from databases (Bind, IntAct, etc.), literature (PubMed), and gene expression resources (such as GEO and ArrayExpress). We illustrate the effectiveness of our approach by asking real-world biomedical questions that link these resources to prioritize the plausible biomarker candidates. Among the 13 prioritized candidate genes, we identified MIF to be a potential emerging candidate due to its role as a pro-inflammatory cytokine. We additionally report on the effort and challenges faced during generation of such an indication-specific knowledge base comprising of curated and quality-controlled data.

Conclusion: Although many alternative approaches have been proposed and practiced for modeling diseases, the semantic web technology is a flexible and well established solution for harmonized aggregation. The benefit of this work, to use high quality and context specific data, becomes apparent in speculating previously unattended biomarker candidates around a well-known mechanism, further leveraged for experimental investigations.

Keywords: RDF, Semantic web, Data integration, Data curation, Data harmonization, Disease modeling, Neurodegenerative diseases, Alzheimer's disease

* Correspondence: shweta.bagewadi@scai.fraunhofer.de
†Equal contributors
[1]Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany
[2]Bonn-Aachen International Center for Information Technology, Rheinische Friedrich-Wilhelms-Universität Bonn, 53113 Bonn, Germany
Full list of author information is available at the end of the article

BioMed Central

© 2016 Kawalia et al. **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.

Background

Alzheimer's disease (AD), the most prominent neurodegenerative disease (NDD), has become a global threat to the aging society, affecting nearly 115 million people by 2050 [1]. The imperfect understanding of the AD etiology has created a large gap in translating the preclinical findings into clinical trials dominantly observed in high drug attrition rates [2]. Early diagnosis and preventive interventions could facilitate substantial reduction in the number of affected cases to 9 million by 2050 [3, 4]. Particularly, reliable biological markers of disease and disease progression could assist in early diagnosis and treatments catered to the patient [5]. In this direction, considerable global research efforts have been dedicated to investigate molecular players underlying AD pathogenic events, contributing to an ever-growing wealth of disparate data. Refinement of this information into actionable knowledge representations requires a good interoperable and formalized framework, capable of inferring potential biomarkers across different facets of the molecular physiology. Additionally, in silico disease models that integrate complementary data from various resources are capable of recapitulating key mechanisms for a given condition [6–8].

Among others, most widely used data integration strategies include data warehousing (e. g., Pathway Commons [9]), data centralization (e. g., UniProt [10], IntAct [11]), and federated databases (e. g., BioMart [12]). An example of a data integration framework is tranSMART [13], which consists of a data warehouse covering various types of data and related data mining applications required for translational research and biomarker discovery workflows. Such a harmonized aggregation of heterogeneous data sources facilitates interpretation over a large knowledge space [14].

However, one fundamental challenge with most of these integration approaches is to cope with the variability and heterogeneity in content, language, and formats of incoming data from different source repositories. Moreover, regular updates of these data resources are necessary to keep up with newly added information and to avoid incompleteness. The inaccessibility to the integrated data resources, due to altered database structure or change in the naming conventions is unavoidable [15]. Semantic web technologies have overcome the above described challenges up to an extent by revolutionizing the lossless exchange of data and formalizing the data format into a computable knowledge [16], calling it “smart data" [17]. The capability of using rich formal descriptions for data and its standardized mapping allows complex querying in a more efficient way without information loss.

Resource Description Framework (RDF) is the World Wide Web Consortium (W3C) proposed standard for semantic integration and modeling of data. RDF uses the syntax of Extensible Markup Language (XML) and imposes structural constraints to represent the meta-data as a set of triples containing directed edges. One big advantage lies in the usage of common namespaces across the different data domains encoded as Unified Resource Identifiers (URIs). Initiatives such as Identifiers.org [18] provide persistent official identifiers in the biomedical domain, allowing sustained interlinking between distinct data resources. This allows high levels of seamless interoperability between data sources and the capability to access and map against additional related data unambiguously, called data federation. On the contrary, large efforts are still needed during an initial definition of the ontologies to build the schema for data representation.

Semantics in life sciences

The idea of semantic web prevails in various domains, including life sciences. Recently, "The Monarch Initiative" [19] has taken the semantic route to enable reasoning over genotype-phenotype equivalence within and across species. They leverage on ontologies to link external curated data resources for generating new hypothesis and prioritizing candidates/variants based on the phenotypic similarity. Stevens et al. [20] launched TAMBIS, multi-data application tool, which allows biologists to formulate complex molecular biology questions to databases such as Swiss-Prot [21], Enzyme [22], CATH [23], BLAST [24], and Prosite [25] through well-defined semantics.

Among the early users of RDF, Lindemann et al. [26] applied it to centralize and flexibly access the heterogeneous and varying quality of medical data obtained from several clinical partners. The importance of semantic mining in the life science domain was brought to limelight by the Bio2RDF project [27], which demonstrated the possibility of querying life science knowledgebases by linking public bioinformatics databases and providing public SPARQL endpoints. Subsequently, Linking Open Drug Data (LODD) [16] demonstrated linking drug data information from DrugBank [28] and clinical trials resources. Chem2Bio2RDF [29] demonstrates the potential usage of the above two mentioned RDF repositories in the field of chemoinformatics.

Observing the immense advantage of linked open data, several major publicly available life science databases such as UniProt, DisGeNet [30], Protein Data Bank Japan (PDBj) [17], and EBI resources such as Gene Expression Atlas [31], ChEMBL [32], BioModels [33], Reactome [34], and BioSamples [35], have made their data available in the form of RDF. Thus, the RDF platform has been increasingly adopted as a standard for data exchange. Amidst prime users of RDF in elucidating disease pathophysiology, Shin et al. [36] demonstrated systematic querying of linked experimental data to

explore the effect of genes that are regulated by volatile organic compounds in human blood. Qu et al. [6] showed semantic framework capability in drug repurposing by proposing Tamoxifen, an FDA approved drug for Breast Cancer, as a candidate drug for Systemic Lupus Erythematosus. The above reported association has already been tested in mice by Sthoeger et al. [37], showing a leverage of semantic web in a real world scenario. Furthermore, Willighagen et al. [38] presented the linkage of several RDF technologies in molecular chemoinformatics and proteochemometrics.

To our knowledge, there has been very limited application of semantic web approaches to the research of neurodegenerative diseases. Linked Brain Data (LBD) [39] is an upcoming initiative which focusses on understanding the brain functionality by integrating resources such as genomic, proteomic, anatomical and biochemical resources with respect to neuroscience. Using such a multi-level knowledgebase, they aim to understand the association between cognitive functions and brain diseases. Lam et al. [40] made the first attempt to develop an e-Neuroscience data integration framework, AlzPharm [41]. They extracted AD-related drug information from BrainPharm [42] to be further integrated with manually inferred hypotheses from the scientific literature and published articles (SWAN [43]). They demonstrated the usage of such a model by clustering AD drugs based on their molecular targets and to filter publications (claims and hypotheses) specific to Donepezil effect on treatment of AD. Although AlzPharm made use of manually inferred hypothesis, they lack the validation of their findings with experimental data such as gene expression and pathways.

Motivation

Despite the current advancements in semantic web technology, we still do not have cure for complex diseases like AD. One of the key reasons accountable for this could be the increasing gap between generated data and the derived knowledge. In order to increase the probability of the derived knowledge to be novel, data quality and data reliability is highly desirable. Moreover, the contextual specificity of the data is of paramount importance.

Compared to relational database management system (RDBMS) technologies, in RDF the relations have explicit meaning (expressiveness) in a given context and are directly accessible; allowing the user to extract meaningful knowledge from the data as opposed to an unstated structured data. In addition, RDF structures are more adaptive and flexible, allowing fluidity in the data relationships. This overcomes the fragility of RDBMS; where if the underlying representation of the keys and flat table changes, the tentacled connections are lost. Moreover, triples from RDF can be transformed into RDBMS structure and vice-versa. One another advantage of RDF is its graph representation that enables us to better explore relations through network topological characteristics such as relatedness, network perturbation, centrality, influence, etc. The usage of automated reasoners have largely been beneficial to understand the semantics and to expand the associated relations [44].

In this paper we propose *NeuroRDF*, an approach harnessing the potential of RDF as a framework for modeling neurodegenerative diseases to enable a close, biologically sensitive integration of well curated, complementary, and multi-faceted data. The fundamental principle of this strategy is to take advantage of semantics to develop a context specific, multi-layered in silico disease model, represented as a formalized, and computationally processable domain knowledge. A fine-grained analysis of the metadata from various data resources empowers the user to ask more focused questions around a hypothesized pathomechanism involving previously neglected or hidden candidates, further leveraged for experimental investigations. Considerable efforts have been invested to process and manually curate huge amounts of data that is required to build such a knowledge base around a specific indication. This includes for example the in-depth assessment of the respective phenotype, the type of tissue used in an experiment, and information around the donor of the tissue like gender, age, and possible comorbidities. Querying such a highly curated and focused knowledgebase increases the chances of unraveling novel hypothesis, which could have been lost over time or pave way to newly emerging knowledge.

We used SPARQL to traverse each of these knowledge graphs (derived from distinct resources) in an integrative manner, allowing highly disease specific analysis of the underlying data. Using this approach, we demonstrate an example on how to prioritize novel candidates in AD mechanism.

Methods

The developed generic semantic web-based workflow integrating heterogeneous data resources is outlined in Fig. 1. This multi-layered model integrates data from various public resources such as databases, literature, and gene expression information. The harmonization of heterogeneous data to build RDF models was achieved by using several data/file parsers. The workflow also includes a pre-processing step to monitor the quality of each incoming data type for specificity.

Data collection and resources

This subsection depicts briefly the different data resources integrated into the *NeuroRDF*.

Database-derived interactions for healthy brain

A closer look into the healthy human brain interactions could improve identification of the dysregulated

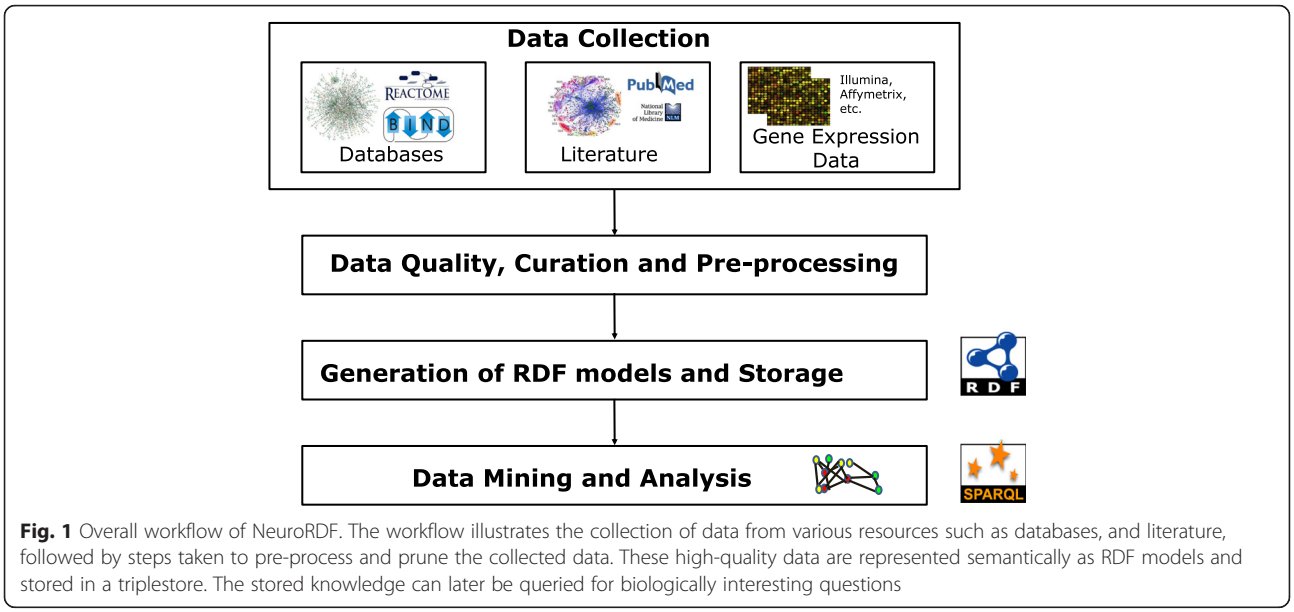


Fig. 1 Overall workflow of NeuroRDF. The workflow illustrates the collection of data from various resources such as databases, and literature, followed by steps taken to pre-process and prune the collected data. These high-quality data are represented semantically as RDF models and stored in a triplestore. The stored knowledge can later be queried for biologically interesting questions

mechanisms which further surges the plausibility of identifying AD drugs in clinical trails [45, 46]. However, the mainstream AD research is biased towards the well known disrupted events such as APP, and tau rather than recognizing their role in normal brain functions [47].

Several publicly available databases provide protein-protein interactions (PPIs) and microRNA-target interactions (MTIs), which can be derived using multiple sources and methodologies. For instance, Human Protein Reference Database [48], Molecular INTeraction database [49], and miRTarBase [50] focus on experimentally verified interactions that are manually mined from literature by expert biologists. In addition to literature-derived information, Biomolecular Interaction Network Database [51] centralizes interactions from high-throughput technologies. Few other databases such as STRING [52], and miRWalk [53] also provide predicted interactions. However, none of these databases mine interactions specific to a given context (for example AD pathology or normal physiology).

A lot of published healthy state PPIs are not directly measured in human cells but in artificial conditions such as human cell lines, human genes transfected into yeast cells, etc., missing out on the biological plausibility in humans and context specificity [54]. Hence, considerable effort by Bossi and Lehner [55] was invested to verify the tissue specificity of PPI interactions from 21 databases (including a few above mentioned) using human gene expression data. Furthermore, this additional action to ensure validity of the interactions in normal state aids improved prediction of genes in disease state [56]. In that direction, our group has extracted a subset of these experimentally confirmed PPIs belonging to healthy brain physiology [57]. Currently, the healthy brain PPI network contains 7,192 genes and 45,001 PPIs.

Extracting AD-specific interactions from literature

The bridging factor between researchers and scientific accomplishments are published as texts, warehoused in large repositories like PubMed [58]. These biomedical articles are the major information source of functional factors such as proteins, genes, microRNAs (miRNAs), etc. However, their functional descriptions are scattered as unstructured text in literature [59]. Text-mining methods could help us mine these articles and retrieve the associated relations/evidence for a given context. Since proteins are the chief players in almost all biological processes and miRNAs have been established in the last decade as important regulators of gene expression, we focus our current research on MTIs and PPIs.

In order to harvest AD-specific knowledge from the literature, we used our in-house state-of-the-art named entity recognition (NER) system ProMiner [60] and the semantic search engine SCAIView [61]. Identification of genes/proteins and disease mentions was accomplished using dictionaries. The disease dictionary was built using MeSH [62], MedDRA [63], and Allie [64] databases. Currently, it contains 4,729 concepts and 64,776 synonyms [65], which are normalized to MeSH names. Human Genes/Proteins dictionary [60] was compiled from three different resources: SwissProt, EntrezGene [66], and HGNC [67]. Currently, this dictionary consists of 36,312 entries and 515,191 synonyms. All the identified gene/protein names were normalized to HUGO gene symbols for maintaining homogeneity across all

data resources and also for future comparisons and visualizations.

To identify MTIs from MEDLINE abstracts, we applied our previously developed approach [65]. Here we extracted novel miRNA mentions using a regular expression. These mentions were normalized to miRBase database identifiers [68]. In addition, relation dictionary containing the major classes of relationship terms between miRNAs and their target genes/proteins was also developed. A tri-occurrence based approach was used to extract the MTIs (co-occurring with a relation term) at the sentence level.

Using the above-mentioned dictionaries, our group previously harvested AD specific PPIs from MEDLINE abstracts and full text articles [69]. Here we used the interaction terms compiled by Thomas et al. [70]. A state-of-the-art machine learning based approach [71] was applied to retain true pairs of PPIs in a given sentence. Both approaches have been optimized for recall. Hence, the obtained relations have been manually filtered for false positives. After manual inspection, 339 PPIs for 301 proteins and 99 MTIs for 36 microRNAs that are specific to AD were obtained. Articles published in languages other than English could lead to increased information content, however a dedicated approach to harvest them is needed. Moreover, separate parsers are needed. Thus, for this work we extracted interactions from the biomedical literature in English.

AD gene expression data

A standard approach to test any generated hypothesis is to assess the gene expression of the involved candidates between affected and healthy patients or in the absence of human data we fall back to animal models or derived cell cultures [72–75]. High-throughput technologies such as microarray, RNA-seq provide potential to measure gene expression simultaneously for different experimental/biological conditions. These studies are assembled in widely adopted public archives: The NCBI Gene Expression Omnibus (GEO) [76] and ArrayExpress [77].

For querying AD-specific gene expression data, we used previously developed database, NeuroTransDB [78], which contains highly curated meta-data information for eligible AD studies. It assembles studies from public resources namely, GEO and ArrayExpress, using a keyword based search approach. Among the 45 prioritized AD human studies, we filtered for microarray studies that measure gene expression in brain tissue extracted from both AD and healthy patients. In addition, availability of raw data was a mandate for applying uniform pre-processing. In total, we obtained eight microarray studies to be integrated in *NeuroRDF*: GSE12685, GSE1297, GSE28146, GSE5281, E-MEXP-2280, GSE44768, GSE44770, and GSE44771.

To assess the quality of the arrays we applied ArrayQualityMetrics [79] package. The selected studies (independent of the platform type) were pre-processed using Bioconductor (Version 3.0) packages in R [80], by applying similar methods for maintaining consistency by reducing variance. All studies conducted on Affymetrix chips were normalized by robust multi-array average method (*rma*) [81]. Similarly, package *limma* [82] was applied on Rosetta/Merck Human 44 k 1.1 microarray chip. All the chips were normalized for background correction and quantile normalization. The normalized intensity values were log2-transformed and duplicate probes were averaged. To identify the differentially expressed genes between healthy and Alzheimer's patients we used *limma* package by applying Benjamini and Hochberg's method to control for false discovery rate (adjusted p-value ≤ 0.05).

Data curation

Although the current text-mining methods have started to leverage expert curators to extract PPIs, MTIs, etc. from text, the extracted information are still prone to false positives [83]. Moreover, it is not straightforward to use these systems for retrieval of context-specific triples due to technological limitations [84]. Hence, the meticulousness of the identified triples to occur in a certain cell type, disease state, or events captured in AD-specific documents is not guaranteed. Thus, the need for manual verification is unavoidable, especially when considering the full text articles. The previously published test corpus used for evaluating the constructed AD PPI network contained AD-specific PPIs extracted by the machine learning approach from 200 full text articles [69]. Manual inspection by the authors resulted in retaining PPIs from 38 articles that are truly specific to AD, thus discarding 81 % of the originally retrieved articles. Similarly, we retained only 68 abstracts from 250 articles (27 %) that were retrieved using our tri-occurrence based approach for AD MTIs [65]. Thus, we can conclude that only about 20–30 % of the (relation extraction based) extracted PPIs and MTIs are truly relevant to AD, pointing out the need of manual curation.

Similarly, in our recent publication [78], we have highlighted the key issues related to retrieval and reusability of the datasets from public transcriptomics archives, such as GEO and ArrayExpress. We showed that a simple keyword based search not necessarily asserts the specificity of the retrieved datasets to the queried disease or organism. When manually inspected, we reported nearly 20 % of these retrieved studies to be irrelevant for AD query. In addition, basic metadata annotations such as age, gender, etc., which strongly contributes to the differential estimates, were observed to be incomplete. Brazma et al. [85] had earlier reported that not all the data submitted to GEO or ArrayExpress are MIAME compliant [86]. We additionally noticed these missing annotations being scattered as unstructured prose in database webpages,

publications, supplementary material, figures, etc., leading to a steep increase in the needed curation effort. Although the published research articles are rich in annotations, a large number of experiments have missing citations [87], which have to be added manually. Moreover, inconsistencies between the information stored in the archives and in the associated publications were also noted. On an average, about 30 min to 2 h of curation effort was needed to retrieve pertinent information for a single dataset. The outcome of this work resulted in a highly curated metadata database, NeuroTransDB, which is used in this work for extracting relevant AD gene expression studies.

Generation of RDF models

RDF data model

RDF allows the generation of models for processed data that exchanges information on the Web [82]. The RDF data model stores all the relationships between different entities as triples (subject-predicate-object). In RDF terminology, the subject, the predicate and the object are known as resources and are represented by a unique "Uniform Resource Identifiers (URIs)" in order to support global data exchange. Literals are constant values such as numbers and strings mapped to the resources. Literals can only be used as objects but never as subjects or predicates.

RDF schemas

We constructed the RDF schemata by abiding the standard RDF graph notation where an ellipse represents Resource, an arrow for Property, and rectangle for Literal. In all the RDF schemas, we have maintained a common resource representation for the "Gene" namespace adapted from Bio2RDF that maps to the NCBI gene database. For the namespaces with no available ontologies, we created an internal namespace, called "SCAI". Some of the properties were described using URIs from Dubin Core Metadata Element [88].

Four separate schemas (for each data resource) have been generated that are centered on genes for interoperability, associating each gene product to its official gene symbol. In the AD PPI schema (see Fig. 2), proteins and their interactions were represented using the Uniprot Core Ontology [89]. Supporting literature evidence were adapted to URIs from Bio2RDF namespaces. The article resource was linked to its PubMed ID, sentence in which the interaction has been mentioned, and the associated journal. Experimental evidence that validates the given interaction (if any) were mapped to BioPax [90], MGED [91], ONTOAD [92], and SCAI namespaces. In the MTI models (see Fig. 3), literature, genes, and proteins namespaces were adapted similarly to the PPIs. To represent the miRNAs, we applied the Bio2RDF namespace that links it to miRBase database [93].

For the PPI schema encoding the healthy state, as seen in Fig. 4, we used the same ontologies as in case of AD PPI. Additional interaction evidence such as brain region, reference database, experimental evidence, and literature information were described using Core, BioPax, and Bio2RDF namespaces.

The microarray schema has two branches that are linked to the experiment: sample details and differential expression analysis. The majority of the resources and properties are linked to URIs from EBI's Atlas (atlas) [94] and MGED [91] namespaces, cf. Fig. 5. Gene expression experiments could contain several samples that are measured in different conditions. A detailed description of

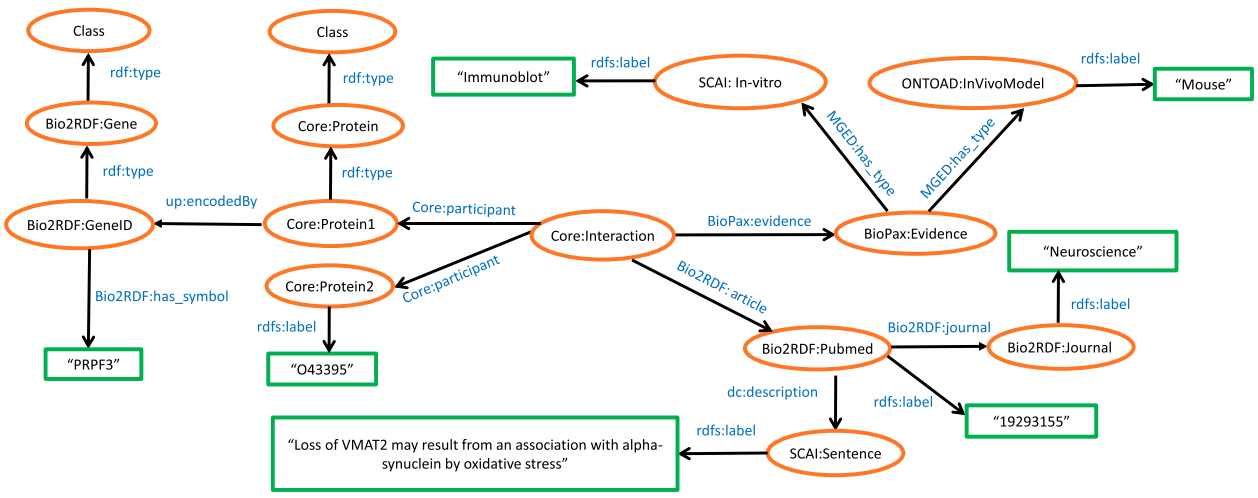


Fig. 2 Schematic representation of the Diseased PPIs in RDF. The figure describes AD specific PPI interactions along with supporting evidence mined from literature

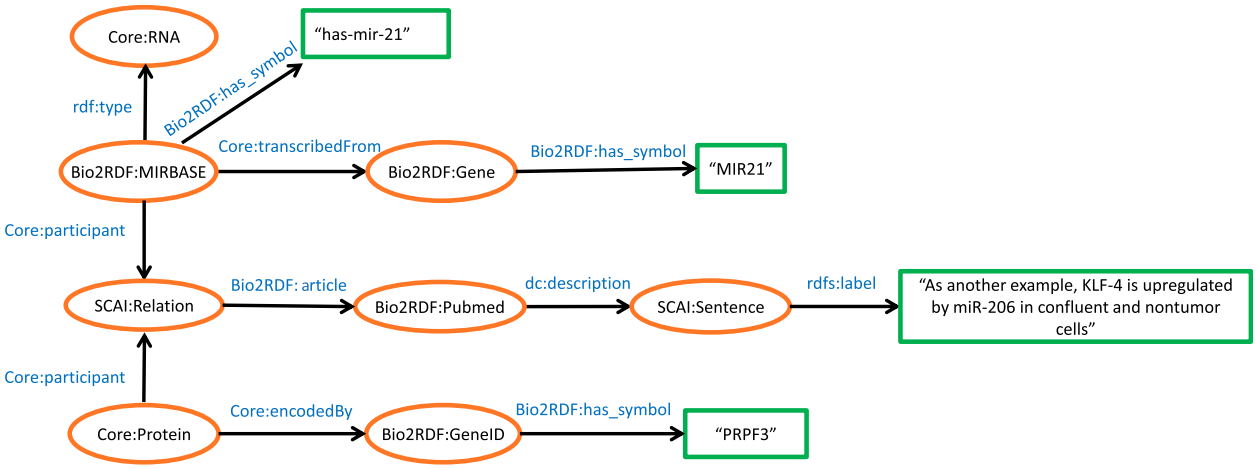


Fig. 3 Schematic representation of MiRNA-target interactions in RDF. The figure encapsulates miRNA mentions along with their corresponding gene identifier from literature

each sample is needed for accurate analysis. Thus, we associated each sample to its meta-data annotations, namely age, gender, organism, organism part, platform, and phenotype. Organism under investigation is mapped to NCBI Taxonomy URIs [95]. The factor value of each sample, i.e., the phenotype information, is described using the EFO ontology [96]. Each platform array is made up of multiple probes that may represent a gene. To be able to retain the expression values for individual probes, we linked the probe ID resource to platform. However, for better reasoning, quantitative values retrieved from statistical analysis are linked to genes and not to probes. The meta-analysis results, derived from *limma* [82], such as differential expression value of a gene and its associated p-value are all linked to the gene symbols.

Construction, validation and storage of RDF models

We modeled all the triples (represented in the schemas) using the Apache Jena API [97]. Resources, and Properties as Java classes were created from the ontologies using the corresponding in-built methods in the API and with the help of Schemagen [98].

In order to check for the correctness of our generated RDF models, we made use of the online service RDF validator [99]. By using such a service, we verified the models using their graph and triples representation.

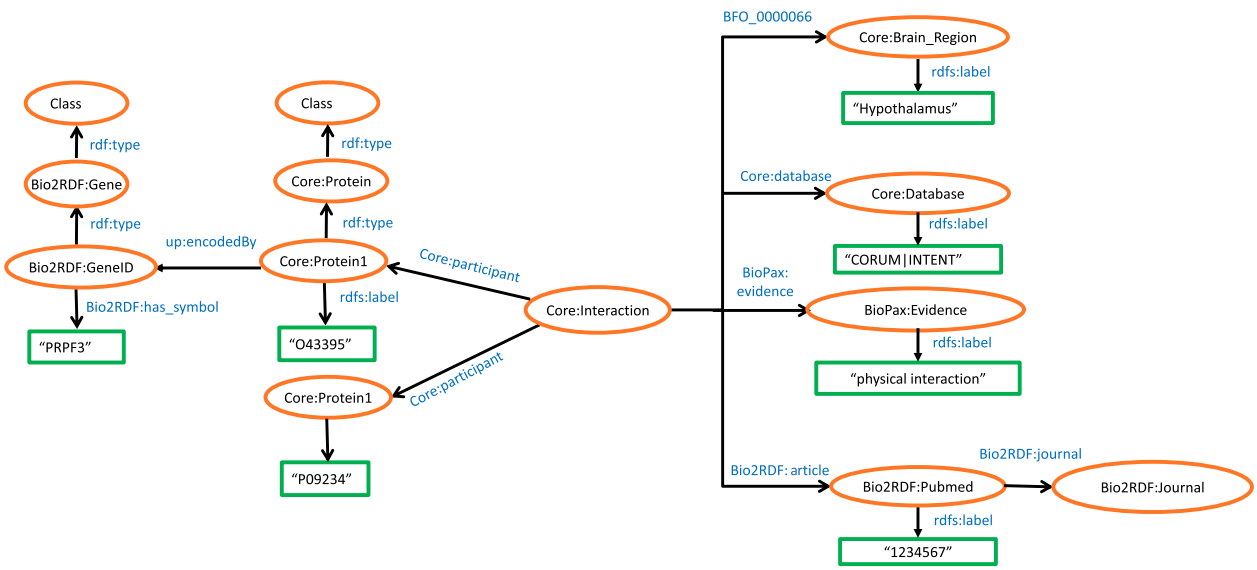


Fig. 4 Schematic representation of Healthy PPIs in RDF. The figure represents PPIs of healthy subjects extracted from literature and PPI specific databases. The schema also contains meta-information about these PPIs

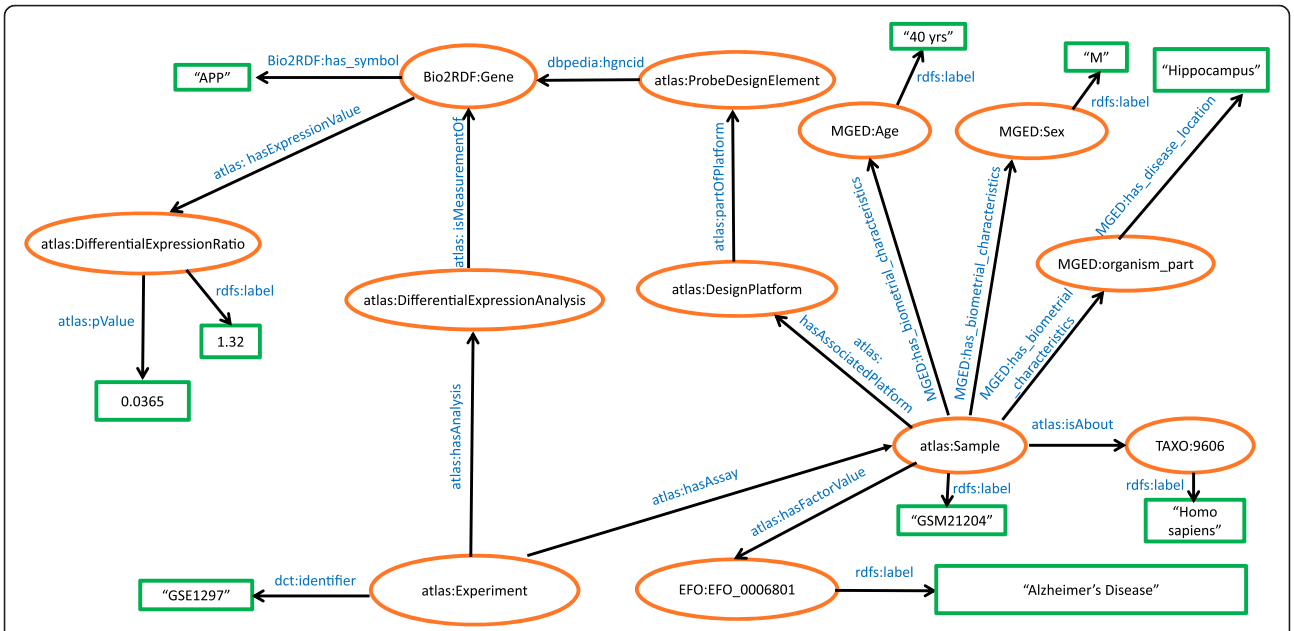


Fig. 5 Schematic representation of Gene Expression Data in RDF. This figure represents gene expression data obtained from public resources such as GEO and ArrayExpress

Triple stores, such as Virtuoso [100], provides an opportunity to store individual or integrated RDF models in one endpoint. Taking advantage of this, we stored all the generated RDF models as individual graphs in a single Virtuoso instance. Using common URIs (e.g., "Gene" identifier) as the connecting link between these models, it is possible to traverse through them integratively.

Data mining and analysis

In RDF, all the stored triples are accessible using a common query language, SPARQL Protocol and RDF Query Language (SPARQL) [101]. We generated a Java library with embedded SPARQL queries to ask our endpoint and the underlying networks biologically relevant questions. Queries were generated from individual models, which were further integrated as nested queries to traverse different graphs. Each query uses the common Gene URI namespace (which is common across all models) to pass on the results used to the next nested query. One possibility to visualize the query results is the SemScape Cytoscape [102], to represent the return values as (sub-) graphs again.

Results and discussions

NeuroRDF covers a wide range of curated AD related data resources, stored as four separate RDF models in a single Virtuoso endpoint. It tries to address the main concepts (complementary) that contributes significantly to unraveling AD pathology.

Differentially expressed genes

For the eight selected microarray datasets, gene expression analysis was performed between healthy and diseased patients. Among these, GSE1297, GSE28146, and E-MEXP-2280 resulted in no differential genes for adjusted p-value cutoff 0.05. From the remaining studies, only genes that exhibited a log2 fold change of > 1.5 were selected for analysis. In total, GSE5281 resulted in 4,278 genes under p-value cutoff and 2 up-, and 48 down-regulated genes for the defined fold change cutoff. Similarly, GSE44770 provided 254 differentially expressed genes, among which 16 up- and 11 down-regulated were selected further. In case of GSE44771, we obtained 335 differential genes that contain 11 up and 11 down-regulated genes that show > 1.5 log2 fold change. For both, GSE12685 and GSE44768, we obtained 1 and 51 genes under the p-value cut-off. However, there were no genes that had log2 fold change of >1.5. The list of all the differentially expressed genes that were selected for further analysis is provided in Additional file 1.

RDF models

Table 1 summarizes the content of the generated triple store by providing some statistics of all integrated networks. In total, there are 8353 unique triples in AD PPI, 1,204,194, 667 unique triples in Healthy PPI, and 20,454 unique triples in gene expression RDF models (Additional file 2). The number of unique predicates (relations) for AD and healthy PPIs are 11, whereas for MTI there are 5 and the gene expression model

Table 1 Statistics of generated RDF models stored in Virtuoso endpoint

Models	No. of triples	No. of entries	No. of properties	Size (mb)
Alzheimer's disease PPI	8353	19900	11	0.894
Healthy State PPI	1204194	78852	11	99.102
MTI	667	300	5	0.095
Microarray	20454	9477	16	303.5

consists of 16. The number of entities present in these models range from 300 to 78,852 (cf. Table 1). In case of the gene expression data, to avoid large triples we excluded the gene expression values of individual probes and included information only related to differential expression. Uploading and querying these models was not computationally expensive due to lower set of predicates and relatively small file size.

Prioritization of AD candidates

To illustrate the potential of NeuroRDF approach and to determine novel AD candidates from the high quality integrated data, we exploit the underlying biological association between the different data resources and identify the previously unknown information.

Our prioritization criteria was based on the notion that every data resource brings with it a piece of missing biological information which is needed to understand the mechanism of a certain candidate. We tried to associate this distributed information by systematically addressing the following questions:

(1) Whether candidates in the diseased network tend to be associated with normal physiology. If yes, what are the common players that could help us in the differential estimates (called as causal candidates);
(2) Which microRNAs regulate the selected causal candidates that could give insights into their post-transcriptional dysregulation;
(3) Have any of the selected causal candidates assessed for their level of differential expression in an unbiased data source (e. g., gene expression data);
(4) How strong is the influence of the neighboring genes on the casual candidates. This is based on the assumption that strong candidates tend be surrounded by dysregulated genes and have an influence on the candidate itself;
(5) Is there any functional relatedness between the causal candidates and their neighbors;

To answer these questions, we generated a set of SPARQL queries. Figure 6 is an example SPARQL query syntax used to obtain miRNAs that regulate the genes in the AD networks. Similar querying has been applied to build a system of faceted searches for the above described questions. Firstly, we identified common genes between the healthy and AD PPI networks. This query resulted in 230 intersecting genes. Looking into the MTIs, we found 13 of these genes to be regulated by at least one microRNA (cf. Table 2). Among these 13 genes, 9 were observed to be differentially expressed: APP, BACE1, ADAM10, IL1B, MAPK3, DLG4, LRP1 PTGS2, and TGFB1. Except for APLP2, and IL6, all the other genes contained differentially expressed neighbors either in AD or in healthy PPIs. There were no miRNAs that were common to these 13 genes.

Sub-networks from the AD and healthy PPIs were extracted to investigate the prioritized candidates (see Figs. 7 and 8). As observed from Fig. 8, for healthy PPIs there was one larger sub-network (containing APP, ADAM10, BACE1, MIF, MAPT, and LRP1) and a smaller one containing two genes (PTGS2, and IL1B). On the other hand, for diseased PPIs in Fig. 7, there were two large sub-networks containing four (STAT4, JUN, MAPK3, and STMN2) and five genes (APP, LRP1, BACE1, DLG4, and TGFB1). The third sub-network was made up of two genes (MAPT, and TUBA4A). Among the prioritized candidates, APLP2 and IL6 had no common links to other prioritized candidates. Thus, they were discarded for further analysis.

Relevance of prioritized AD candidates

The remarkability of complementing wet lab research using the predictability and reproducibility of measured outcomes is one of the core reasons why researchers are

```
SELECT ?Gene ?Rel ?Mirna ?Gene2
    from <http://localhost:8890/MiRNA>
    where {
        ?Gene <http://purl.uniprot.org/core/encodedBy> ?Protein .
        ?Protein <http://purl.uniprot.org/core/participant> ?Rel .
        ?Mirna <http://purl.uniprot.org/core/participant> ?Rel .
}
```

Fig. 6 Example SPARQL query for information retrieval from NeuroRDF. SPARQL query as seen in the figure retrieves the miRNAs for a given gene

Table 2 Prioritized AD candidate genes

Intersected genes between healthy and AD PPI	MiRNAs	Differentially expressed neighbors		Number of literature articles for intersected genes
		Healthy PPI	AD PPI	
APP	MIR101-1, MIR106A, MIR106B, MIR124-1, MIR137, MIR153-1, MIR181-C, MIR29A, MIR520C, MIR19-1	ADAM10, MAPT, MIF, BACE1, LRP1	TGFB1, BACE1, LRP1	24550
BACE1	MIR107, MIR124-1, MIR145, MIR298, MIR29A, MIR29B1, MIR328, MIR9-1	APP	APP, LRP1	1883
ADAM10	MIR451, MIR144, MIR1306, MIR107, MIR103	APP	-	231
IL1B	MIR146A, MIR155	PTGS2	-	1099
MAPK3	MIR15A, MIR155	-	STMN2, JUN	276
MAPT	MIR16-1, MIR132	APP	TUBA4A	3367
APLP2	MIR153-1	-	-	134
DLG4	MIR485	-	LRP1	151
IL6	MIR27B	-	-	748
JUN	MIR144	-	STAT4, MAPK3	142
LRP1	MIR205	APP	DLG4, APP, BACE1	305
PTGS2	MIR146A	IL1B	-	474
TGFB1	MIR155	-	APP	276

This table summarizes the literature based evidences of intersected genes between healthy and AD PPI and their corresponding miRNAs and differentially expressed genes

more inclined to the field of bioinformatics. Therefore, in silico validation of predicted candidates for its relevancy is of utmost importance. In this direction, we pinpoint the relevance of our prioritized candidates through a literature survey.

AD established candidates

Although there are no FDA approved biomarkers for AD, researchers focus on some of the key candidates that are hypothesized to be involved in AD. In the current NDD research practice, APP has been established as the widely used biomarker candidate. The classical pathological hallmark of AD is formation of amyloid-beta aggregates (leading to plaques) in brain. This is reported to be caused by faulty proteolytic processing of APP that releases amyloid-beta [103]. Another hallmark of AD is tau pathology (MAPT gene), regulated by amyloid-beta. Hyperphosphorylation of tau causes accumulation of neurofibrillary tangles due to the disrupted functioning of axonal transport [5]. However, it is also interesting to note the paradigm shift in AD research due to recently failed drug trails that focused mostly around these hypotheses [2]. Nevertheless, several neuroscientists still believe in the potential of APP and the tau hypothesis for elucidation of the underlying pathomechanism. As observed from our generated sub-networks, our largest sub-network was established around the APP gene.

When compared to APP, BACE1 has not been so frequently studied. However these genes often fall into the "most interesting gene zone" as far as AD is concerned since it is involved in the formation of amyloid-beta. BACE1 is the major enzyme (beta secretase) involved in the cleaving of APP at beta site and generating soluble amyloid-beta [104]. However, increased BACE1 activity has been reported to be associated with amyloid-beta aggregation in AD patients [105]. Bu et al. have detailed out the evidence that LRP1 is a receptor for APOE, a contributing factor to AD [106]. Furthermore, in 1993, Strittmatter, Roses and colleagues [107] have identified APOE4 as the major risk for late-onset AD. TGFB1 polymorphism has been widely associated with an increased risk of late-onset AD. Deficiency in TGFB1 signaling leads to neurofibrillary tangle formation increasing the advancement of mild cognitive impairment patients to AD, by increasing the depressive symptoms [108]. DLG4 is a post-synaptic scaffolding protein that interacts with postsynaptic receptors such as NMDA receptors for efficient postsynaptic response [109]. However, its impairment has largely contributed to the synaptic degeneration in AD. Mutations in ADAM10 gene have been associated to late-onset AD. ADAM10 enzyme has alpha-secretase activity to cleave amyloid-beta, however BACE1 competes with ADAM10 for cleavage. Thus, its decreased expression has been implicated in AD pathogenesis [110].

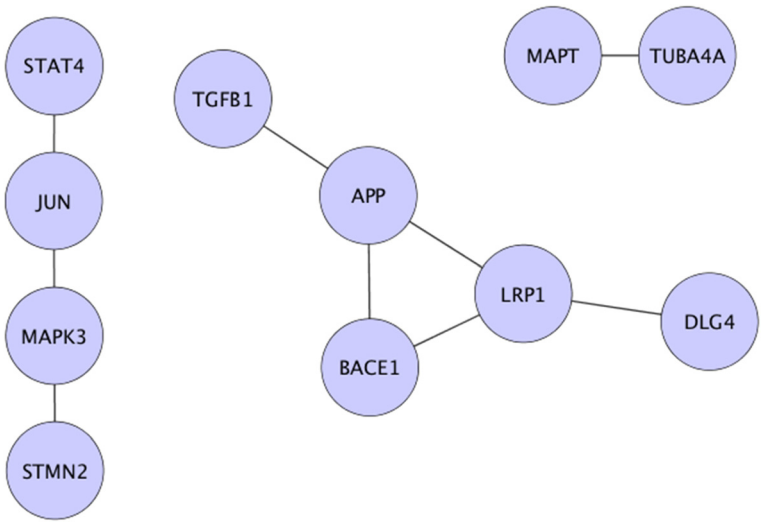


Fig. 7 Extracted sub-networks from AD PPIs network. This figure symbolizes the diseased sub-graphs that were generated using prioritized candidates and their differentially expressed neighbors

AD emerging candidates

To identify emerging knowledge in the context of AD, we performed an individual gene analysis using SCAIView for publications in PubMed. Here, we measured the co-occurrence of the causal genes (including its differential neighbors) and AD over a period of last 10 years, see Fig. 9. Since the number of articles for the APP gene was relatively too high each year, we normalized the number of literature evidence of other candidates using the APP gene's article count for that year. Hence, the normalized range for the literature distribution is between 0 and 1, where 1 is the highest number of articles for that year (the APP gene). Please refer to Additional file 3 for details of the literature counts. Inspecting literature evidence, we found that all the prioritized causal candidates have been studied in the context of AD. Moreover, among their differentially expressed neighbors, STMN2 (8 articles), MAPK4 (1 article), TUBA4A (2 articles), and MIF (15 articles) contained fewer articles related to AD. Among these genes, STMN2 and MIF have been recently studied in the context of AD, whereas, MAPK4, STMN2, and TUBA4A were implicated in AD nearly two decades before but failed to establish as robust biomarker candidates.

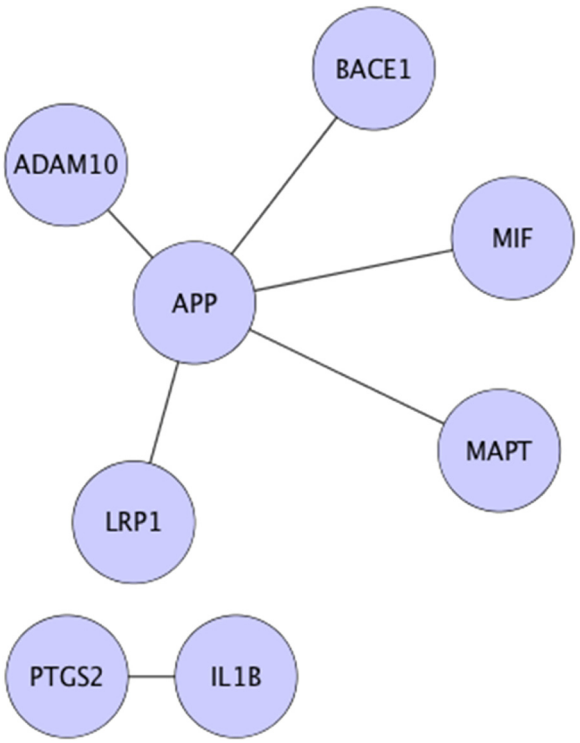


Fig. 8 Extracted sub-networks from healthy PPIs network. This figure symbolizes the healthy sub-graphs that were generated using prioritized candidates and their differentially expressed neighbors

MIF's role in AD

Macrophage Migration Inhibitory Factor (MIF) has for long been known to participate in tumor proliferation due to its pro-inflammatory cytokine functionality [111]. In general, MIF acts as a key regulator of inflammatory activities such as innate and adaptive immunity [112]. Apart from that, it is also known to play a significant role as an anti-apoptotic factor of neutrophils as well as macrophages [113].

The MIF gene has been well studied in cancer and inflammation. However, recent studies are emerging around a plausible role of MIF in neurodegenerative diseases, in particular AD. Moreover, Flex et al. [114] have earlier reported that MIF polymorphisms are not linked

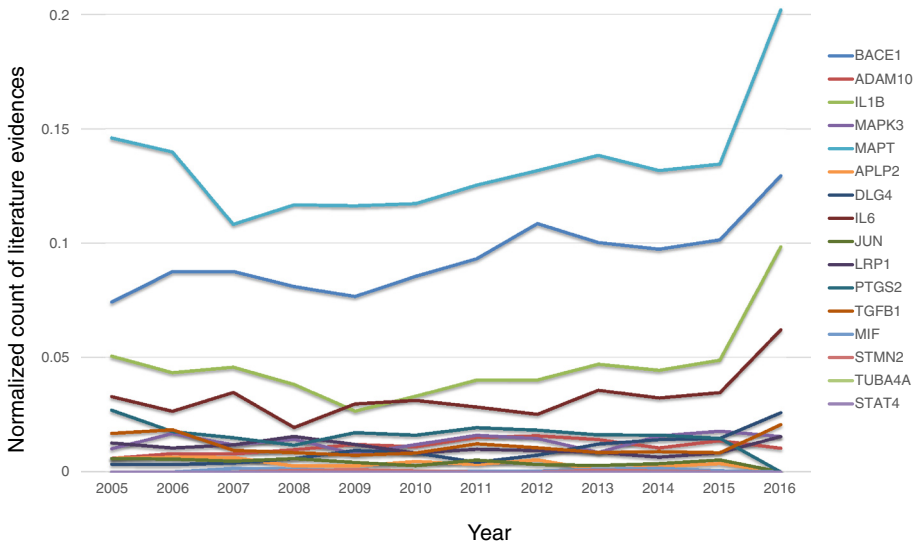


Fig. 9 Statistics of the literature evidence for emerging candidate genes in the speculated sub-networks

to AD, but confirmed its complex immune and inflammatory activities. Although, APP and tau have been associated to play a key role in the pathophysiology of AD, many researchers strongly believe in the role of inflammatory processes subsidizing to the pathology of AD. This stems from the fact that activated microglial cells discharge immunoregulatory cytokines which result in various side-effects such as neuronal dysfunction and inhibition of hippocampal neurogenesis [115]. MIF is one such pro-inflammatory cytokine which is known to bind with amyloid-beta protein and enhance the plaque removal and neuronal debris from the brain during normal conditions [116]. Also, MIF has been identified to play a role in neuronal survival by inhibiting the activation of ERK-1/MAP kinases [117] (regulatory role in cell proliferation and glucocorticoid action) as well as its ability to surpass the p53 mediated apoptosis [118]. Although, the precise molecular function of MIF in the context of AD is unknown, it is known to play a role in inflammatory processes around the plaque formation. MIF is also highly expressed in the neurons of rat hippocampus, one of the primary regions to be affected by AD [117]. Bryan et al. [119] also report on the abnormal expression of MIF in both microglia and in the hippocampal neurons in human. This all makes MIF a plausible biomarker for inflammatory responses in AD.

Conclusion

NeuroRDF approach has been designed to identify new knowledge through semantic mining. The proposed integrative approach takes advantage of the RDF technology to integrate well-curated data from various sources within a specific indication area. From our perspective, it is necessary to focus on one indication or at least a group of indications to build such a knowledge base for precise modeling and analysis due to the high curation effort one has to spend in order to reach the necessary details. We showed how to harmonize three major heterogeneous resources (databases, gene expression data, and literature) used in the research area to generate hypotheses for underlying disease mechanisms. This approach supports identification of novel insights without compromising over quality. Furthermore, new data resources can be included without altering the overall framework. The usage of well-accepted ontologies provides the advantage for further integration of external resources and databases (e.g., federated queries). Using such an approach, we were able to prioritize MIF gene as an emerging candidate due to its role in inflammatory processes implicated in AD pathogenesis.

The advantage of using an RDF schema is that it is highly supportive for data interoperability. Although this work represents the usage of the RDF schema specific for AD, we have also extended the same to other disease models such as Parkinson's and Epilepsy. However, the curated data and the generated hypothesis for these two diseases will be released in future under the terms of a Neuroallianz agreement [120]. Also, these resources are constantly kept up-to-date as they are transferred to various upcoming projects such as AETIONOMY [121].

Additional files

Additional file 1: List of differentially expressed genes. This file contains the list of differentially expressed genes (for each dataset used) that fall under the adjusted p-value cutoff of 0.05. The differential expression analysis was performed using *limma* package in R statistical environment. The file is provided in an Excel format. (XLSX 68 kb)

Additional file 2: The developed RDF models and the SPARQL queries used are made available at: http://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/downloads/neurordf.html. (ZIP 178 kb)

Additional file 3: Detailed count of literature evidences for prioritized candidates. This file contains the detailed count of number of evidences available for each prioritized candidate for each year since 2005 in context of Alzheimer's disease. These statistics were retrieved using SCAIView knowledge discovery tool (as of 18 May, 2016). (XLSX 35 kb)

Acknowledgement
We are grateful to Matthew Page, Translational Bioinformatics, UCB Pharma for providing his valuable inputs during the design of the project and reviewing the manuscript. We are thankful to Erfan Younesi and Ashutosh Malhotra for providing the healthy state PPI and AD-PPI network respectively for this work. We also want to thank Christian Ebeling for his support in building the resources for gene expression data. We would like to acknowledge the Semantic Mining in Biomedicine (SMBM2014) conference organizers, participants, and reviewers for inspiring discussions during the conference. The authors express gratitude to the SMBM2014 conference organizers for providing an opportunity to submit to the Journal of Biomedical Semantics an extended version of the initially published conference proceeding paper.

Funding
This study was funded by a grant from the German Federal Ministry for Education and Research (BMBF) within the BioPharma initiative "Neuroallianz".

Authors' contributions
AI, SBK, PS, and MHA conceived and designed the overall research strategy required for data integration. PS is the scientific supervisor to this work. SBK and AI are the main contributors to manuscript writing. TR contributed to the analysis of gene expression data. PS, and MHA reviewed the content. All authors read and approved the final version of the manuscript.

Competing interests
The authors declare that they have no competing interests.

Declarations
The underlying principles of this article have been previously published in Proceedings of the 6th International Symposium on Semantic Mining in Biomedicine (SMBM2014), Aveiro, Portugal, 2014.

Author details
[1]Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany. [2]Bonn-Aachen International Center for Information Technology, Rheinische Friedrich-Wilhelms-Universität Bonn, 53113 Bonn, Germany. [3]University of Applied Sciences Koblenz, RheinAhrCampus, Joseph-Rovan-Allee 2, 53424 Remagen, Germany.

Received: 1 March 2015 Accepted: 23 May 2016
Published online: 08 July 2016

References

1. International AD. Policy brief for heads of government: the global impact of dementia 2013–2050. 2013. http://www.alz.co.uk/research/G8-policy-brief.
2. Golde TE, Schneider LS, Koo EH. Anti-Aβ therapeutics in Alzheimer's disease: the need for a paradigm shift. Neuron. 2011;69:203–13. doi:10.1016/j.neuron.2011.01.002.
3. Brookmeyer R, Johnson E, Ziegler-Graham K, et al. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement. 2007;3:186–91. doi:10.1016/j.jalz.2007.04.381.
4. Norton S, Matthews FE, Barnes DE, et al. Potential for primary prevention of Alzheimer's disease: An analysis of population-based data. Lancet Neurol. 2014;13:788–94. doi:10.1016/S1474-4422(14)70136-X.
5. Rachakonda V, Pan TH, Le WD. Biomarkers of neurodegenerative disorders: how good are they? Cell Res. 2004;14:347–58. doi:10.1038/sj.cr.7290235.
6. Qu XA, Gudivada RC, Jegga AG, et al. Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships. BMC Bioinf. 2009;10 Suppl 5:S4. doi:10.1186/1471-2105-10-S5-S4.
7. Le Masson G, Przedborski S, Abbott LF. A computational model of motor neuron degeneration. Neuron. 2014;83:1–14. doi:10.1016/j.neuron.2014.07.001.
8. Talwar P, Silla Y, Grover S, et al. Genomic convergence and network analysis approach to identify candidate genes in Alzheimer's disease. BMC Genomics. 2014;15:199. doi:10.1186/1471-2164-15-199.
9. Pathway Commons Database. http://www.pathwaycommons.org/about/. This and all the subsequent URLs have been accessed on 31 May 2016.
10. UniProt Database. http://www.uniprot.org/
11. IntAct Database. http://www.ebi.ac.uk/intact/
12. BioMart. http://www.biomart.org/
13. Szalma S, Koka V, Khasanova T, et al. Effective knowledge management in translational medicine. J Transl Med. 2010;8:68. doi:10.1186/1479-5876-8-68.
14. Rodriguez-Esteban R, Loging WT. Quantifying the complexity of medical research. Bioinformatics. 2013;29:2918–24. doi:10.1093/bioinformatics/btt505.
15. Aoki-Kinoshita KF, Kinjo AR, Morita M, et al. Implementation of linked data in the life sciences at BioHackathon 2011. J Biomed Semantics. 2015;6:3. doi:10.1186/2041-1480-6-3.
16. Samwald M, Jentzsch A, Bouton C, et al. Linked Open drug data for pharmaceutical research and development. J Cheminform. 2011;3:19. doi:10.1186/1758-2946-3-19.
17. Kinjo AR, Suzuki H, Yamashita R, et al. Protein Data Bank Japan (PDBj): Maintaining a structural data archive and resource description framework format. Nucleic Acids Res. 2012;40:453–60. doi:10.1093/nar/gkr811.
18. Identifiers.org. http://identifiers.org
19. The Monarch Initiative. http://monarchinitiative.org/page/about
20. Stevens R, Baker P, Bechhofer S, et al. TAMBIS: transparent access to multiple bioinformatics information sources. Bioinformatics. 2000;16:184–5. doi:10.1147/sj.402.0532.
21. Swiss-Prot Database. http://web.expasy.org/docs/
22. Enzyme Database. http://enzyme.expasy.org
23. CATH Database. http://www.cathdb.info
24. BLAST. http://blast.ncbi.nlm.nih.gov/Blast.cgi
25. Prosite Database. http://prosite.expasy.org
26. Lindemann G, Schmidt D, Schrader T, et al. The resource description framework (RDF) as a modern structure for medical data. Int J Biol Life Sci. 2008;4:89–92. http://waset.org/publications/3109/the-resource-description-framework-rdf-as-a-modern-structure-for-medical-data.
27. Belleau F, Nolin MA, Tourigny N, et al. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. J Biomed Inform. 2008;41:706–16. doi:10.1016/j.jbi.2008.03.004.
28. DrugBank Database. http://www.drugbank.ca
29. Chen B, Dong X, Jiao D, et al. Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. BMC Bioinf. 2010;11:255. doi:10.1186/1471-2105-11-255.
30. Furlong LI. DisGeNET : from MySQL to nanopublication, modelling gene-disease associations for the semantic Web. Paris: Proc 5th Int Work Semant Web Appl Tools Life Sci; 2012. *Fr Novemb 28–30, 2012 2012* Published Online First: 2012. http://ceur-ws.org/Vol-952.
31. Kapushesky M, Adamusiak T, Burdett T, et al. Gene Expression Atlas update–a value-added database of microarray and sequencing-based functional genomics experiments. Nucleic Acids Res. 2012;40:D1077–81. doi:10.1093/nar/gkr913.
32. CHEMBL Database. https://www.ebi.ac.uk/chembl/
33. BioModels Database. http://www.ebi.ac.uk/biomodels-main/
34. Reactome Ontology. http://www.reactome.org
35. BioSamples Database. http://www.ebi.ac.uk/biosamples/
36. Shin GH, Kang YK, Lee SH, et al. MRNA-centric semantic modeling for finding molecular signature of trace chemical in human blood. Mol Cell Toxicol. 2012;8:35–41. doi:10.1007/s13273-012-0005-9.
37. Sthoeger ZM, Zinger H, Mozes E. Beneficial effects of the anti-oestrogen tamoxifen on systemic lupus erythematosus of (NZBxNZW)F1 female mice are associated with specific reduction of IgG3 autoantibodies. Ann Rheum Dis. 2003;62:341–6. doi:10.1136/ard.62.4.341.
38. Willighagen EL, Alvarsson J, Andersson A, et al. Linking the resource description framework to cheminformatics and proteochemometrics. J Biomed Semantics. 2011;2 Suppl 1:S6. doi:10.1186/2041-1480-2-S1-S6.
39. Linked Brain Data. http://www.linked-neuron-data.org/about.jsp?link=link6
40. Lam HYK, Marenco L, Clark T, et al. Semantic Web Meets e-Neuroscience : An RDF Use Case, Semant Web - ASWC 2006 first Asian semant web conference. 2006. p. 158–70.
41. Lam HYK, Marenco L, Clark T, et al. AlzPharm: integration of neurodegeneration data using RDF. BMC Bioinf. 2007;8 Suppl 3:S4. doi:10.1186/1471-2105-8-S3-S4.

42. BrainPharm Database. http://senselab.med.yale.edu/BrainPharm
43. SWAN Ontology. http://www.w3.org/TR/hcls-swan
44. Hoehndorf R, Schofield PN, Gkoutos GV. The role of ontologies in biological and biomedical research: a functional perspective. Brief Bioinform. 2015;16:1069–80. doi:10.1093/bib/bbv011.
45. Douaud G, Refsum H, de Jager CA, et al. Preventing Alzheimer's disease-related gray matter atrophy by B-vitamin treatment. Proc Natl Acad Sci. 2013;110:9523–8. doi:10.1073/pnas.1301816110.
46. Tagawa K, Homma H, Saito A, et al. Comprehensive phosphoproteome analysis unravels the core signaling network that initiates the earliest synapse pathology in preclinical Alzheimer's disease brain. Hum Mol Genet. 2015;24:540–58. doi:10.1093/hmg/ddu475.
47. Kodamullil AT, Younesi E, Naz M, et al. Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis. Alzheimer's Dement. 2015;11:1329–39. doi:10.1016/j.jalz.2015.02.006.
48. Human Protein Reference Database (HPRD). http://www.hprd.org/
49. The Molecular INTeraction Database (MINT). http://mint.bio.uniroma2.it/mint/Welcome.do
50. Chou CH, Chang NW, Shrestha S, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. 2015;5712121:gkv1258. doi:10.1093/nar/gkv1258.
51. Biomolecular Interaction Network Database (BIND). http://bioinformatics.ca/links_directory/database/9267/bind-biomolecular-interaction-network-database
52. STRING Database. http://string-db.org/
53. miRWalk Database. http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/
54. Schaefer MH, Lopes TJS, Mah N, et al. Adding protein context to the human protein-protein interaction network to reveal meaningful interactions. PLoS Comput Biol. 2013;9, e1002860. doi:10.1371/journal.pcbi.1002860.
55. Bossi A, Lehner B. Tissue specificity and the human protein interaction network. Mol Syst Biol. 2009;5:260. doi:10.1038/msb.2009.17.
56. Magger O, Waldman YY, Ruppin E, et al. Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. PLoS Comput Biol. 2012;8, e1002690. doi:10.1371/journal.pcbi.1002690.
57. Younesi E, Hofmann-Apitius M. Biomarker-guided translation of brain imaging into disease pathway models. Sci Rep. 2013;3:3375. doi:10.1038/srep03375.
58. PubMed Database. http://www.ncbi.nlm.nih.gov/pubmed
59. Krallinger M, Erhardt RA, et al. Text mining approaches in molecular biology and biomedicine. Drug Discov Today. 2005;10:439–45.
60. Fluck J, Mevissen HT, Dach H, et al. ProMiner: recognition of human gene and protein names using regularly updated dictionaries, Proceedings second BioCreative challenge evaluation work. Madrid: CNIO; 2007. p. 149–51.
61. SCAIView. http://www.scaiview.com/en/scaiview-distributions/scaiview-academia.html
62. National Library of Medicine's MeSH Controlled Vocabulary. http://www.ncbi.nlm.nih.gov/mesh
63. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf. 1999;20:109–17. doi:10.2165/00002018-199920020-00002.
64. Allie Database. http://allie.dbcls.jp/
65. Bagewadi S, Bobić T, Hofmann-Apitius M, et al. Detecting miRNA mentions and relations in biomedical literature. F1000 Res. 2014; doi: 10.12688/f1000research.4591.2
66. NCBI's Entrez Gene Database. http://www.ncbi.nlm.nih.gov/gene
67. HUGO Gene Nomenclature Committee (HGNC). http://www.genenames.org/
68. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42:D68–73. doi:10.1093/nar/gkt1181.
69. Malhotra A, Younesi E, Sahadevan S, et al. Exploring novel mechanistic insights in Alzheimer's disease by assessing reliability of protein interactions. Sci Rep. 2015;5:13634. doi:10.1038/srep13634.
70. Thomas P, Solt I, Klinger R, et al. Learning to extract protein – protein interactions using distant supervision. In: Proceedings of robust unsupervised and semi-supervised methods in natural language processing, Workshop at international conference recent advances in natural language processing. 2012.
71. Bobić T, Klinger R, Thomas P, et al. Improving distantly supervised extraction of drug-drug and protein-protein interactions, Proc 13th Conf Eur Chapter Assoc Comput Linguist. 2012. p. 35–43.
72. Kogelman LJA, Cirera S, Zhernakova DV, et al. Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. BMC Med Genomics. 2014;7:57. doi:10.1186/1755-8794-7-57.
73. Krämer A, Green J, Pollard J, et al. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics. 2014;30:523–30. doi:10.1093/bioinformatics/btt703.
74. Van Dam D, De Deyn PP. Animal models in the drug discovery pipeline for Alzheimer's disease. Br J Pharmacol. 2011;164:1285–300. doi:10.1111/j.1476-5381.2011.01299.x.
75. McDermott JE, Wang J, Mitchell H, et al. Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. Expert Opin Med Diagn. 2012;7:1–15. doi:10.1517/17530059.2012.718329.
76. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10. doi:10.1093/nar/30.1.207.
77. Brazma A, Parkinson H, Sarkans U, et al. ArrayExpress - A public repository for microarray gene expression data at the EBI. Nucleic Acids Res. 2003;31:68–71. doi:10.1093/nar/gkg091.
78. Bagewadi S, Adhikari S, Dhrangadhariya A, et al. NeuroTransDB : highly curated and structured transcriptomic metadata for neurodegenerative diseases. Database. 2015;2015:bav099. doi:10.1093/database/bav099.
79. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics - A bioconductor package for quality assessment of microarray data. Bioinformatics. 2009;25:415–6. doi:10.1093/bioinformatics/btn647.
80. Bioconductor. http://www.bioconductor.org/
81. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4:249–64. doi:10.1093/biostatistics/4.2.249.
82. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47. doi:10.1093/nar/gkv007.
83. Czarnecki J, Shepherd AJ. Mining biological networks from full-text articles. Methods Mol Biol. 2014;1159:135–45. doi:10.1007/978-1-4939-0709-0_8.
84. Krallinger M, Vazquez M, Leitner F, et al. The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text. BMC Bioinf. 2011;12:S3. doi:10.1186/1471-2105-12-S8-S3.
85. Brazma A. Minimum Information About a Microarray Experiment (MIAME)–successes, failures, challenges. Sci World J. 2009;9:420–3. doi:10.1100/tsw.2009.57.
86. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet. 2001;29:365–71. doi:10.1038/ng1201-365.
87. Piwowar H, Chapman W. Recall and bias of retrieving gene expression microarray datasets through PubMed identifiers. J Biomed Discov Collab. 2010;5:7–20. doi:10.5210%2Fdisco.v5i0.2785.
88. Dublin Core Metadata Element Set. http://dublincore.org/documents/dces/
89. Uniprot Core Ontology. http://lov.okfn.org/dataset/lov/vocabs/uniprot
90. Biological Pathway Exchange (BioPax). http://www.biopax.org/
91. Whetzel PL, Parkinson H, Causton HC, et al. The MGED ontology: a resource for semantics-based description of microarray experiments. Bioinformatics. 2006;22:866–73. doi:10.1093/bioinformatics/btl005.
92. Ontology of Alzheimer's Diseases and Related Diseases (ONTOAD). http://bioportal.bioontology.org/ontologies/ONTOAD
93. The miRBase Database. http://www.mirbase.org/.
94. Atlas RDF Ontology. https://www.ebi.ac.uk/fgpt/ontologies/gxaterms.html.
95. NCBI Taxonomy Namespace. http://www.ncbi.nlm.nih.gov/taxonomy.
96. Malone J, Holloway E, Adamusiak T, et al. Modeling sample variables with an experimental factor ontology. Bioinformatics. 2010;26:1112–8. doi:10.1093/bioinformatics/btq099.
97. Jena Tutorial. https://jena.apache.org
98. Schemagen Documentation. http://jena.apache.org/documentation/tools/schemagen.html
99. RDF Validator. http://www.w3.org/RDF/Validator
100. Virtuoso. http://virtuoso.openlinksw.com
101. Sparql. http://www.w3.org/TR/rdf-sparql-query
102. Cytoscape Tool. http://apps.cytoscape.org/apps/semscape
103. Golde TE, Petrucelli L, Lewis J. Targeting Aβ and tau in Alzheimer's disease, an early interim report. Exp Neurol. 2010;223:252–66. doi:10.1016/j.expneurol.2009.07.035.
104. Cole SL, Vassar R. The Alzheimer's disease beta-secretase enzyme, BACE1. Mol Neurodegener. 2007;2:22. doi:10.1186/1750-1326-2-22.
105. Washington PM, Morffy N, Parsadanian M, et al. Experimental traumatic brain injury induces rapid aggregation and oligomerization of amyloid-beta in an Alzheimer's disease mouse model. J Neurotrauma. 2014;31:125–34. doi:10.1089/neu.2013.3017.

106. Bu G. Apolipoprotein E, and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. Nat Rev Neurosci. 2009;10:333–44. doi:10.1038/nrn2620.
107. Strittmatter WJ, Saunders AM, Schmechel D, et al. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci. 1993;90:1977–81. doi:10.1073/pnas.90.5.1977.
108. Bosco P, Ferri R, Grazia Salluzzo M, et al. Role of the transforming-growth-factor-β1 gene in late-onset Alzheimer's disease: implications for the treatment. Curr Genomics. 2013;14:147–56. doi:10.2174/1389202911314020007.
109. Leuba G, Vernay A, Kraftsik R, et al. Pathological reorganization of NMDA receptors subunits and postsynaptic protein PSD-95 distribution in Alzheimer's disease. Curr Alzheimer Res. 2014;11:86–96. doi:10.2174/1567205013106660170.
110. Vassar R. ADAM10 prodomain mutations cause late-onset Alzheimer's disease: not just the latest FAD. Neuron. 2013;80:250–3. doi:10.1016/j.neuron.2013.09.031.
111. Choi S, Kim H-R, Leng L, et al. Role of macrophage migration inhibitory factor in the regulatory T cell response of tumor-bearing mice. J Immunol. 2012;189:3905–13. doi:10.4049/jimmunol.1102152.
112. Calandra T, Roger T. Macrophage migration inhibitory factor: a regulator of innate immunity. Nat Rev Immunol. 2003;3:791–800. doi:10.1038/nri1200.
113. Baumann R. Macrophage migration inhibitory factor delays apoptosis in neutrophils by inhibiting the mitochondria-dependent death pathway. FASEB J. 2003;17:2221–30. doi:10.1096/fj.03-0110com.
114. Flex A, Pola R, Serricchio M, et al. Polymorphisms of the macrophage inhibitory factor and C-reactive protein genes in subjects with alzheimer’s dementia. Dement Geriatr Cogn Disord. 2004;18:261–4. doi:10.1159/000080026.
115. Dong CJ, Guo Y, Ye Y, et al. Presynaptic inhibition by 2 receptor/adenylate cyclase/PDE4 complex at retinal Rod bipolar synapse. J Neurosci. 2014;34:9432–40. doi:10.1523/JNEUROSCI.0766-14.2014.
116. Oyama R, Yamamoto H, Titani K. Glutamine synthetase, hemoglobin α-chain, and macrophage migration inhibitory factor binding to amyloid β-protein: their identification in rat brain by a novel affinity chromatography and in Alzheimer's disease brain by immunoprecipitation. Biochim Biophys Acta Protein Struct Mol Enzymol. 2000;1479:91–102. doi:10.1016/S0167-4838(00)00057-1.
117. Mitchell RA, Metz CN, Peng T, et al. Sustained Mitogen-activated Protein Kinase (MAPK) and Cytoplasmic Phospholipase A2 Activation by Macrophage Migration Inhibitory Factor (MIF): regulatory role in cell proliferation and glucocorticoid action. J Biol Chem. 1999;274:18100–6. doi:10.1074/jbc.274.25.18100.
118. Mitchell RA, Liao H, Chesney J, et al. Macrophage migration inhibitory factor (MIF) sustains macrophage proinflammatory function by inhibiting p53: Regulatory role in the innate immune response. Proc Natl Acad Sci. 2002;99:345–50. doi:10.1073/pnas.012511599.
119. Bryan KJ, Zhu X, Harris PL, et al. Expression of CD74 is increased in neurofibrillary tangles in Alzheimer's disease. Mol Neurodegener. 2008;3:13. doi:10.1186/1750-1326-3-13.
120. The Neuroallianz Consortium. http://www.neuroallianz.de/en/mission.html
121. Aetionomy. www.aetionomy.eu

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

SUMMARY

Earlier AD research was predominantly focused on the amyloid cascade and tau pathology hypotheses. However, recent developments in clinical trials indicate that more than 200 therapeutic trials targeting the above two assumptions have consistently failed to demonstrate any therapeutic recovery. This goes on to confirm that AD is not caused by a set of molecular dysregulations, but instead, it is a syndrome of multiple factors ranging from cellular to behavioral to lifestyle ones. This requires then a multi-modeling approach, which provides access to all multi-layer data in one platform such that meaningful insights into disease progression are attained.

With the rapid advancements in next-generation sequencing technologies, data warehouses and continually expanding literature resources, the need for efficient data management and retrieval system is an absolute necessity. With the explosion of new data, a simultaneous challenge of assembling and maintaining a comprehensive platform that contains all relevant information also exists. Moreover, the era of digital health has made it inevitable for biologists and medical doctors to be equally competent in handling online technologies. Through data integration, it is now possible to intelligently mine the data and also ask meaningful, context-specific questions to the system to get better insights in addressing scientific problems. Efficient cross-talk between data resources can only be achieved when the data are semantically integrated across all data scales, namely genes, proteins, compounds, pathways, drugs, and diseases.

Semantic web technology is one such framework which facilitates the smooth integration of heterogeneous data resources, ranging from cellular level data to

clinical level data, with the help of ontologies and terminologies. Using RDF data models, or graphical representations of integrated data, it is easily possible to traverse and bridge multiple layers of embedded knowledge and thereby augment the possibility to derive novel insights into AD pathomechanisms. Additionally, semantic web framework offers one practical solution to address the problem of multi-scale modeling as it enables the integration of all quantitative and qualitative data ranging from molecular targets to clinical readouts. This framework is represented in the form of a graphical network, which facilitates network traversing across scales along with the possibility of querying and retrieving relevant information from the system.

The earliest usage of semantic web technology in the life sciences was done to prioritize human candidate genes from a pool of heterogeneous genomic and phenomic databases. Over the years, semantic web technology has been widely exploited for exploring complex biological phenomena, especially in the field of disease modeling such as in cancer and AD.

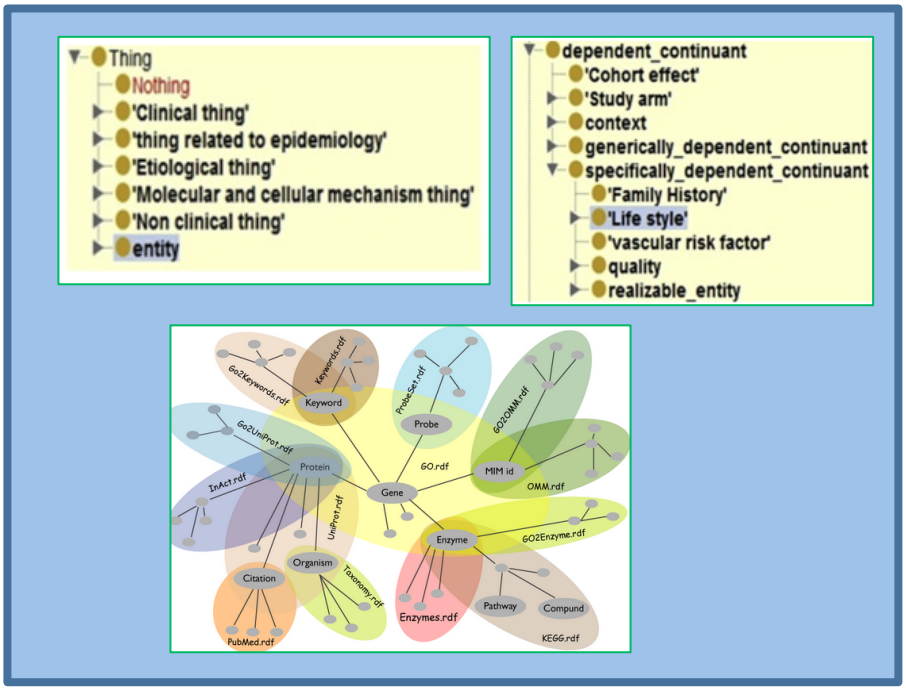
The work presented in this section focuses extensively on state-of-the-art semantic web technologies used for developing contextually rich, heterogeneous models for NDD based diseases. This integrative framework, known as "NeuroRDF", comprises of integrated data from three different domains, namely: Protein-protein interaction (PPI) databases, contextual knowledge from the literature on PPIs and transcriptomics data. The core objective of this work lies in the ability to get novel insights into complex disease mechanisms using an integrative framework through semantic mining. Another advantage of such an integrative framework is that the

existing framework can be reused for analyzing other diseases without altering the overall structure. Additionally, the framework can be easily extended to other data domains such as metabolomics and clinical data.

One of the significant challenges that emerged during the extension of the NeuroRDF model to other data domains was the lack of reliable text-mining tools for harvesting domain-specific metabolomics and clinical data from the literature. These two challenges are addressed in Chapter 2 and 3, respectively.

CHAPTER 2

Knowledge acquisition for elucidating novel dysfunctional pathways through domain specific semantic framework



INTRODUCTION:

The era of systems biology and genome-wide throughput technologies have revolutionized the approach for analyzing biological processes on a systematic level rather than at the level of individual molecular interactions. Pathways are the most preferred representatives of system level interaction information. In particular, molecular and signaling pathways comprise of the interactions of biomolecules that perform specific cellular and molecular functions. These processes are regulated by signals or chemical reactions, which can either enhance or disrupt the biological process, thereby becoming an exciting target for drug therapy and treatment. For this reason, biological pathway databases are often considered the most reliable source of knowledge.

However, the initial step of constructing pathways is often very challenging as it involves extensive manual curation for intrinsic knowledge from the literature. With the exponential growth of scientific publications, it is nearly impossible to conduct manual curation without the assistance of text mining technologies. In spite of significant advancements in biomedical ontology research, the semi-automated process of extracting context-specific, pathway relevant information from the literature remains elusive.

The publication presented in this section focuses on the need for a pathway inventory that can ensure broad coverage of the complete pathway knowledge domain on NDD.

CONTRIBUTION TO THE MANUSCRIPT

My contributions for the following manuscript are:

- I was entirely responsible for building the Pathway Terminology System (PTS) terminology.
- I have performed most of the analysis of the manuscript except for PTS performance evaluation, Fingerprinting of patents.
- I have contributed to the generation of most of the figures and tables in the manuscript
- I have also contributed to the writing of the following sections of the manuscript - Abstract, Materials and Methods, Results and Discussion
- Michaela Gündel contributed to the following sections – PTS system description and evaluation and Discovery of pathway knowledge by fingerprint patents.

Journal of Alzheimer's Disease 52 (2016) 1343–1360
DOI 10.3233/JAD-151178
IOS Press

Towards a Pathway Inventory of the Human Brain for Modeling Disease Mechanisms Underlying Neurodegeneration

Anandhi Iyappan[a,b,1], Michaela Gündel[a,b,1], Mohammad Shahid[c], Jiali Wang[d], Hui Li[e], Heinz-Theodor Mevissen[a], Bernd Müller[f], Juliane Fluck[a], Viktor Jirsa[g], Lia Domide[h], Erfan Younesi[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 Information Technology, Bonn, Germany*
[c]*Population Health Sciences, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany*
[d]*University of Luxembourg, Faculty of Science, Technology and Communication, Life Sciences Research Unit, Luxembourg-Limpertsberg, Luxembourg*
[e]*No affiliation*
[f]*ZBMED Leibniz Information Centre for Life Sciences, Cologne, Germany*
[g]*Aix-Marseille Université, Institut de Neurosciences des Systèmes, Faculté de Médecine, Marseille, France*
[h]*Codemart, Str. Petofi Sandor, Cluj Napoca, Romania*

Accepted 9 March 2016

Abstract. Molecular signaling pathways have been long used to demonstrate interactions among upstream causal molecules and downstream biological effects. They show the signal flow between cell compartments, the majority of which are represented as cartoons. These are often drawn manually by scanning through the literature, which is time-consuming, static, and non-interoperable. Moreover, these pathways are often devoid of context (condition and tissue) and biased toward certain disease conditions. Mining the scientific literature creates new possibilities to retrieve pathway information at higher contextual resolution and specificity. To address this challenge, we have created a pathway terminology system by combining signaling pathways and biological events to ensure a broad coverage of the entire pathway knowledge domain. This terminology was applied to mining biomedical papers and patents about neurodegenerative diseases with focus on Alzheimer's disease. We demonstrate the power of our approach by mapping literature-derived signaling pathways onto their corresponding anatomical regions in the human brain under healthy and Alzheimer's disease states. We demonstrate how this knowledge resource can be used to identify a putative mechanism explaining the mode-of-action of the approved drug Rasagiline, and show how this resource can be used for fingerprinting patents to support the discovery of pathway knowledge for Alzheimer's disease. Finally, we propose that based on next-generation cause-and-effect pathway models, a dedicated inventory of computer-processable pathway models specific to neurodegenerative diseases can be established, which hopefully accelerates context-specific enrichment analysis of experimental data with higher resolution and richer annotations.

Keywords: Alzheimer's disease, disease mechanism, disease modeling, neurodegeneration, pathway terminology

[1]These authors contributed equally to this work.

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

ISSN 1387-2877/16/$35.00 © 2016 – IOS Press and the authors. All rights reserved

INTRODUCTION

"Molecular pathways" represent a series of actions among molecules in a cell. They encompass various biochemical entities, such as enzymes, signaling proteins, and DNA and RNA molecules, which interact to perform certain cellular functions. A pathway can be characterized based on the flow of information (signal) or chemical reactions under a specific condition at a specific time with the aim of regulating cellular biological processes [1]. It is noteworthy that the "pathway" term has been also used in the literature for describing anatomical connections between two brain regions or two cellular organelles, and for functional connectivity or connections (systematic co-expression of physiological activity at two different sites of the brain). This terminology is distinct from molecular signaling pathways.

The role of signaling pathways in health and disease has been massively investigated in the literature and specific databases have been dedicated to categorizing and storing pathway information. Examples include the Kyoto Encyclopedia of Genes and Genomes (KEGG) [2], the BioCarta [3] database, and WikiPathways [4]. Unfortunately, there is little overlap in the pathway content of such databases [5], the contextual information (e.g., condition- and cell-specific information) is often missing, and their content usability is limited by low resolution of interactions and incomplete annotation [6]. Moreover, depending on the actions they perform in the cell, the functional output they regulate or the disease context in which they occur, molecular signaling pathways have been named and classified differently. However, it was until recently that the Pathway Ontology (PW) [7], developed at the Rat Genome Database (RGD) [8], was introduced for representing all types of pathways under a single, standard framework [9]. This ontology provides a valuable resource for annotation of genes, drugs, and diseases to pathways but its application can go beyond simple annotation of terms and entities. Although terms and annotations in the PW have been extracted from the literature, these pieces of information are disconnected from the mechanistic context and actual events underlying various normal and disease conditions. For example, it would be very interesting to mine the literature for signaling pathways that collectively contribute to Alzheimer's disease (AD) pathology and compare their roles to pathways active under the normal, healthy condition. We have recently shown that such a differential analysis strategy can provide invaluable insights into perturbed mechanisms that are unique to a disease state [10].

In fact, reconstruction of pathway involvement in the etiology of neurodegenerative diseases lays the ground for a better understanding of the mechanisms that are involved in the pathophysiology of those diseases. It ultimately bears the potential to provide guidance for drug discovery strategies. Nevertheless, before that, there is an urgent need to delineate both structural and functional complexity of the human brain. Several initiatives including the Neuroscience Information Framework (NIF) [11] and The Virtual Brain (TVB) [12] have undertaken efforts to address structural and functional complexity of the human brain, respectively, by building dedicated resources. Even so, existing gaps between these resources should be bridged and their integrative application should be leveraged so that fundamental questions about the human brain under normal and abnormal conditions can be answered. For example, the lack of an overview on cellular pathways that regulate region-specific functions of the human brain hampers efforts to bridge the gap between structural and functional characteristics of the brain. A few studies have attempted to identify pathways underlying regional and cell type changes in neurodegenerative diseases using differential gene expression profiles (e.g., [13] and [14]); but systematic annotation of pathways to human brain regions using published knowledge has not yet been performed. Compilation and study of a "human brain pathway map" under healthy and diseased conditions will lend support to mechanistic interpretation of disease progression at spatial and temporal resolution.

A mechanistic understanding of disease etiology is a particular challenge in the area of neurodegenerative diseases such as AD and Parkinson's disease (PD), as they are widely known as being "idiopathic", i.e., their etiology is unknown. In order to support the systematic gathering of pathway information linked to brain anatomy and neurodegenerative diseases, we have developed automated approaches that build on dedicated terminologies, tools, and text mining methods for the efficient extraction of pathway information from the scientific literature. In this paper, we present the Pathway Terminology System (PTS), which integrates existing pathway and biological event ontologies, and demonstrate its applications to retrieving context-specific pathway information from the literature. We generate an inventory of pathways active in various regions and cell types of the human brain under healthy and AD

conditions, and analyze drug mode-of-action through mechanistic model of targeted pathway in the context of a large, brain-specific, drug-target-network, the Human Brain Pharmacome (HBP).

MATERIALS AND METHODS

Generation of the Pathway Terminology System (PTS)

Based on the PW, we created a pathway dictionary and integrated terms from the INOH [15] event ontology. Since we re-used events from INOH hierarchy and pathways from PW, their original classification hierarchies as well as their individual concept annotations were preserved after merging these two into PTS. For INOH event classes, we retained their namespace (hereafter abbreviated "IEV") and identifiers, whereas for dictionary classes, we created stable identifiers in their newly generated namespace. As INOH event contains several unclassified upper-level classes, we classified these classes inside the INOH event hierarchy. Since we differentiate between events and pathways, we moved the IEV:Pathway class to be a sibling class of IEV:Event, that, in turn, are both subclasses of SPAN:Processual entity [16] (cf. Fig. 1). We furthermore enriched the pathways with genes involved using an object property pointing from a pathway to its genes. We achieved this enrichment by downloading the necessary information from KEGG using the BioConductor KEGGREST service [17].

Moreover, the merged PTS was populated with pathways from four popular public biomedical pathway databases, namely KEGG, Reactome [18], BioCarta, and Pathway Interaction Database [19]. The pathway names from KEGG and Reactome were mapped to their identifiers using reactome.db and KEGG.db packages in R, which enables an automated updating of the pathway dictionary on the

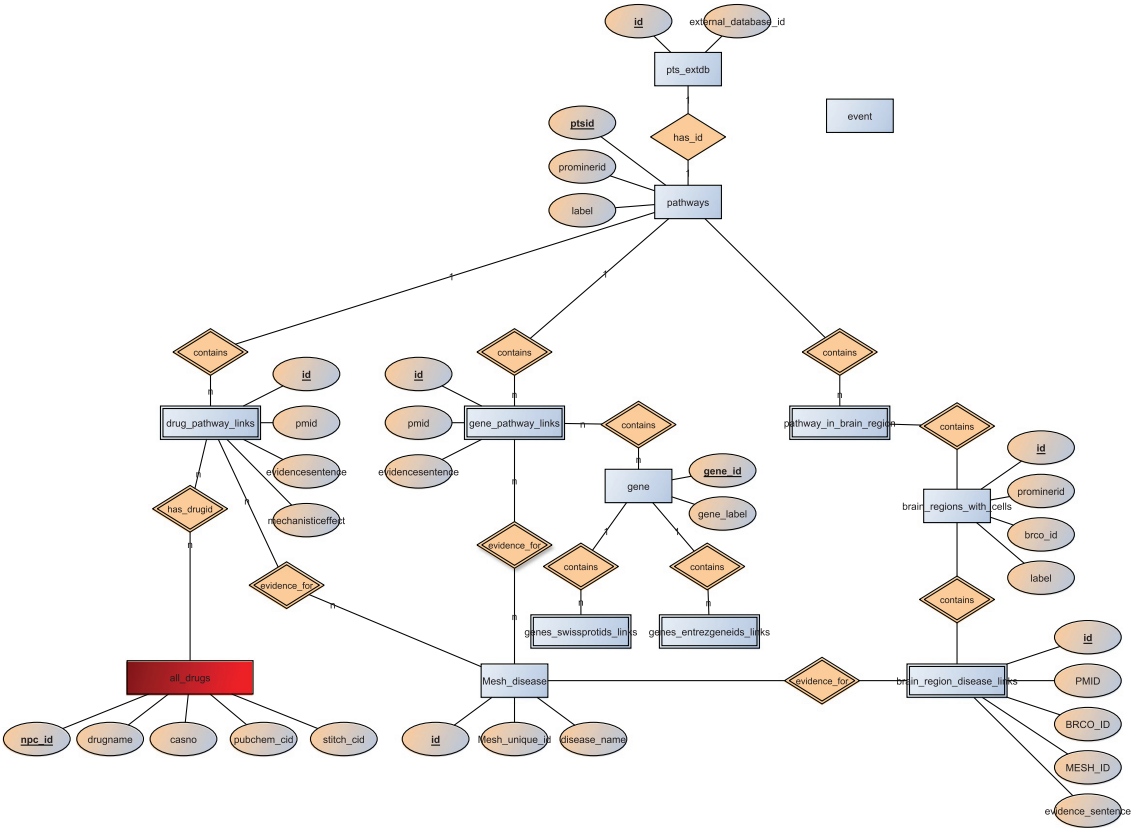


Fig. 1. Overall integration of HBP database with MySQL tables. This figure depicts the integration of pathways, genes, diseases, brain regions and cells into the in-house HBP database, represented in red box. The figure also highlights the MySQL table linking the newly generated information with the existing HBP database.

basis of the pathway identifiers in different pathway databases in the future.

Further enrichment of the PTS was achieved by analysis of relevant scientific text. For this purpose, the phrases (with frequency of occurrence) from MEDLINE abstracts containing the word "pathway" and three words preceding it were extracted using a sub-corpus. If the pathway name inside the four-word phrase was either absent from the pathway reference name list or absent from the pathway synonym list, it was added to the pathway reference name dictionary. Meanwhile, the pathway synonym list was enriched in two steps:

1. Synonym search through the Internet for each term in the pathway dictionary (e.g., Google Scholar).
2. Synonyms detected from a training set of MEDLINE abstracts (300 manually annotated MEDLINE abstracts).

Most of the synonyms were collected by scanning diverse knowledge sources, which include review articles, content of online books, standard knowledge bases, encyclopedias, glossaries, and informative online sources and websites.

Human Brain Pharmacome (HBP)

The HBP is our central knowledge base representing our current knowledge on human brain pharmacology. The HBP integrates data from known drugs and developmental or experimental compounds that are related to human brain disorders, their interacting drug targets, pathways containing these drug targets and the associated phenotypic data and brain regions involved.

The HBP data represents multi-scale biological and chemical knowledge, which has been extracted from open source databases. This dataset has been further enriched by a series of similarity computations performed over drugs and protein 3D structures. Besides, complementary information extracted from biomedical literature, patent documents, and electronic health records were also added to this resource using ProMiner software [20] together with our own pipeline programmed in Perl and Shell code to generate data on co-occurring entities in MEDLINE abstracts. These co-occurrences were then manually curated for validity, transformed into MySQL tables and integrated with the HBP database as depicted in Fig. 1. For instance, co-mentioning of brain drugs with pathway information was used to annotate the HBP with the pathway information related to drug treatments.

A list of MeSH [21] terms for all brain disorders, diseases and disease subtypes was used as a query to search and retrieve data about drugs and their drug targets. The external database resources used include DrugBank [22], Comparative Toxicogenomic Database (CTD) [23], Therapeutic Target Database (TD) [24], EMBL-STITCH Database [25], SIDER [26], ChEMBL [27], BindingDB [28], PDSP KiDB [29], KEGG Database [30], and PhenomicDB [31].

The HBP and the pathway annotation of the HBP will be made publicly available in the course of the IMI project EPAD [32]. Tailored for use in EPAD, the HBP, together with pathway annotations, will become part of the AETIONOMY [33] knowledge base, which in turn will be part of the IT infrastructure underlying EPAD.

Brain Region and Cell type Ontology (BRCO)

BRCO captures a wide range of key concepts representing human brain neuroanatomical structures and integrates their corresponding cell types. BRCO re-uses and connects to multiple existing anatomical ontologies and terminologies, including the NIF Standard ontology [34], the NeuroNames terminology [35], and the Brenda Tissue Ontology (BTO) [36]. The ontology was further enriched with concepts and synonyms extracted from various web-based resources, e.g., Unified Medical Language System (UMLS) [37]. Extensive manual and computational curation was also employed to merge duplicated or similar concepts from different taxonomies and check for ambiguities. BRCO was further tested with regard to formal consistency using the FaCT++ Description Logic Reasoner [38]. From BRCO, we derived a dedicated terminology (BRCT) [39] that could be used to describe and retrieve human brain anatomy concepts from free text. By applying dictionary-based stemming techniques and allowing rule-based lenient match, the final dictionary extracted from BRCO showed a satisfactory F-score of 0.80 in a named entity recognition task using our in-house text mining tool ProMiner on an independent testing corpus composed of 100 manually annotated MEDLINE abstracts. BRCT classes were mapped to pathway and disease mentions (AD, PD, ALS, EP, MS) wherever this contextual link could be extracted from the literature.

Querying in the semantic search environment SCAIView

We integrated the PTS into the literature-mining environment SCAIView [40]. This environment enabled us to perform very context-specific literature searches based on combined semantics from multiple ontologies and terminologies. Such queries are listed in section 2.8. All terminologies and ontologies used in this work are publicly accessible through SCAIView Academia [41].

PTS performance evaluation

We evaluated the performance of PTS based on classical text-mining measures: Recall, precision and F-score. For doing so, we used the following formulas:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$$

$$\text{F-score} = (2 * \text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall})$$

TP, number of true positive hits correctly found, i.e., matching the annotation in the gold standard. FP, number of false positive hits, i.e., hits found but not contained in the gold standard. FN, number of false negative hits, i.e., entities not found but contained in the gold standard. Precision, proportion of correct hits out of all hits. Recall, proportion of correct hits out of all terms that should have been correctly found. F-score, overall measure of accuracy (harmonic mean of precision and recall).

We annotated a set of 300 independent MEDLINE (U.S. National Library of Medicine) abstracts and a small set of 9 full text articles relevant to pathways using the Knowtator tool (BioNLP) [42]. These corpora served as "gold standard" (expected result set) that we subsequently used for evaluation. We also calculated inter-annotator agreement (IAA) using the Knowtator Protégé plugin [43] to determine the consistency and quality of the annotations. During the evaluation, abstracts and full text publications were queried both for the expected "gold standard" named entities and for entities contained in the PTS (class labels and synonyms), using the ProMiner NER tool. These runs were compared and precision, recall and F-score were calculated for the two abstract runs and the two full-text runs, respectively.

Fingerprinting of patents

We used the PTS to fingerprint patents based on pathway terms that occur in the patent text. We used the semantic distances (Wu & Palmer similarity: 2H/(2H + D1 + D2) [44] between pathways to calculate the similarity values of pathway mentions in patents. We chose the Wu & Palmer similarity metric because (i) it is computationally not too expensive and (ii) it calculates similarities by allocating higher similarity values to classes lower down the class hierarchy than to top-level classes. The patent corpus for the fingerprinting approach was generated from PatBase [45]. Using the following formulated query, 4,285 PD relevant patents were retrieved:

(((FT = (Parkinson)) and ((FT = (therapy
OR compound OR formulation OR molecule
OR drug OR agent OR inhibit OR target
OR structure OR treat)) and
(PD = 20070101 : 20121231) and
(IC = (A61P25/16)))),

where FT represents search for the defined terms in the patent's "full text" that are published between 2007 and 2012 (abbreviated as "PD") and are tagged to a specific International Patent Classification (IC) hierarchy. Similarly, 1572 AD relevant patents were retrieved with slight query modification (FT = (Alzheimer) and IC = (A61P25/28)), and patents related to Epilepsy were received via (FT = (epilepsy) or FT = (seizure) and IC = (A61P25/08)).

Among these, 100 patents for each query were manually selected based on high relevancy to the disease. Selected patents were downloaded from PatBase and converted to.xmi format using ChemoCR [46] (for image PDF's) and PDFBox [47] (for text PDFs). Patent fingerprinting and subsequent similarity calculation were implemented using a Java program that runs as a UIMA [48] pipelet. The pipelet uses the input patent.xmi files, starts a ProMiner run on them that annotates the.xmi files with PTS pathway mentions, and creates a further annotation for each patent.xmi file that contains a vector of [0|1]'s (see Fig. 2). These vectors have the length of the sum of all PTS classes (over 6,000), the position in the vector corresponds to the specific PTS class. A "1" indicates occurrence of the respective class (or a synonym) in the patent; a "0" indicates non-occurrences. These vectors are, in a second step, compared to each other and similarities among the patent documents are calculated, resulting in a matrix of similarity values. A value of 1.0 indicates a perfect match, i.e., the same

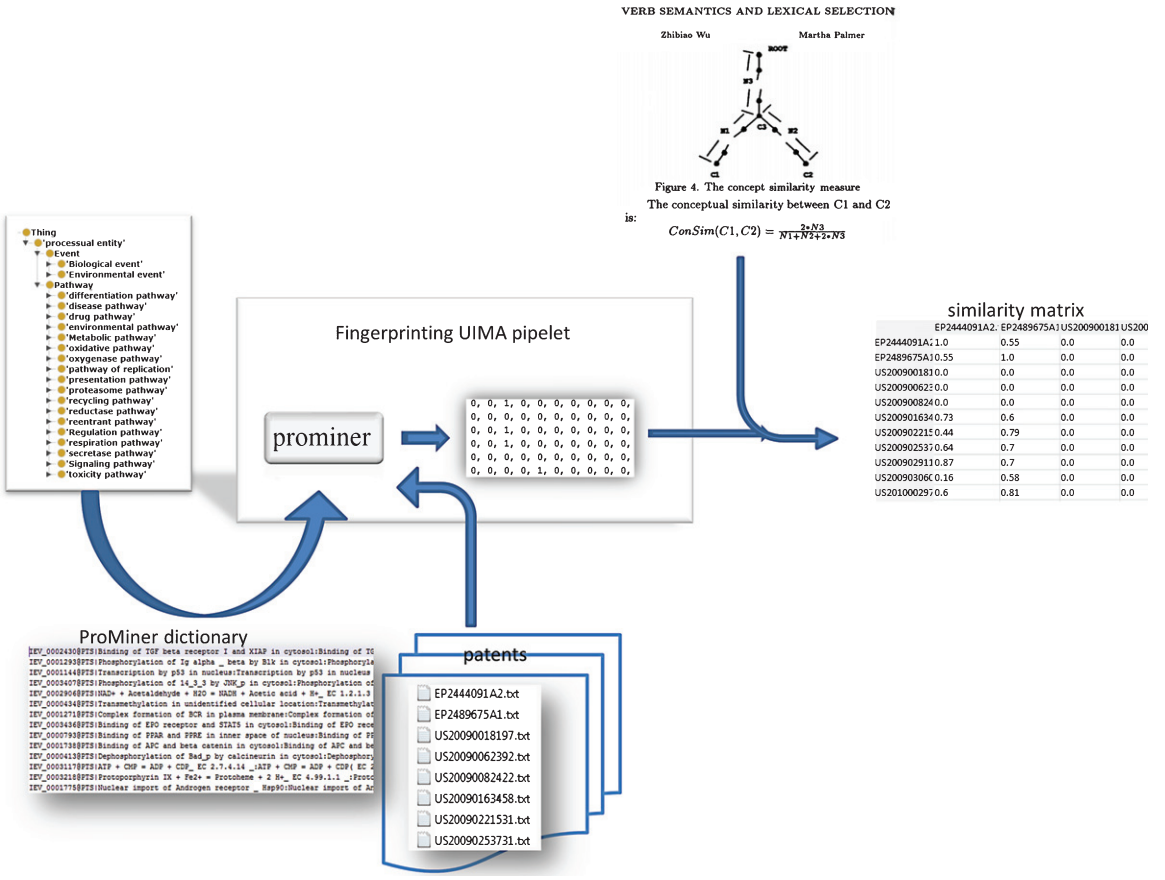


Fig. 2. Patent fingerprinting workflow. The patent fingerprinting works by providing ProMiner dictionary and patents as inputs and the output is a binary vector of 0 and 1, which represents occurrence/non-occurrence of terms.

PTS mentions were found in both patents, a value of 0.0 reflects total dissimilarity of the respective patents. Inter-class similarity is calculated using the Wu & Palmer similarity scoring whereas inter-patent similarity (as each patent may contain $1 \ldots n$ class mentions) is based on the mean value of maximum similarities calculated for the single classes:

$$\text{sim}(\text{patx}, \text{paty}) = \text{mean}\ (\text{sum}\ (\text{max}(\text{sim\_WuPalmer}\ (\text{class\_X\_path\_x\_all\_classes\_pat\_y}))))$$

Based on the similarity scores in the matrix, the most dissimilar patents (i.e. those patents containing pathway mentions other than very well-known pathways involved in AD pathogenesis) were selected (Supplementary File 1). The content of these patents was submitted to the Count-text online tool [49] and the average sentence length was calculated so that the distance of 138 characters was determined for sentence extraction. Within this distance, co-mentions of pathway-gene and pathway-drug were identified and their relationships were manually checked by an expert.

Pathway map of the human brain

To build the pathway map of the human brain, we used a tri-occurrence approach to automatically extract pathways, events and brain region mentions from the literature using PTS and BRCT (Fig. 3). We, then, performed a comparative analysis of pathways between the healthy and AD conditions by finding co-occurrences of pathway mentions with brain regions and cell types in the literature. We targeted the hippocampus region, as it is a critical area for memory and learning affected under AD condition. We made use of the following query for retrieving articles specific to AD: (((([BRCT:“Hippocampus”]) AND [Organism:“Homo sapiens”]) AND [MeSH

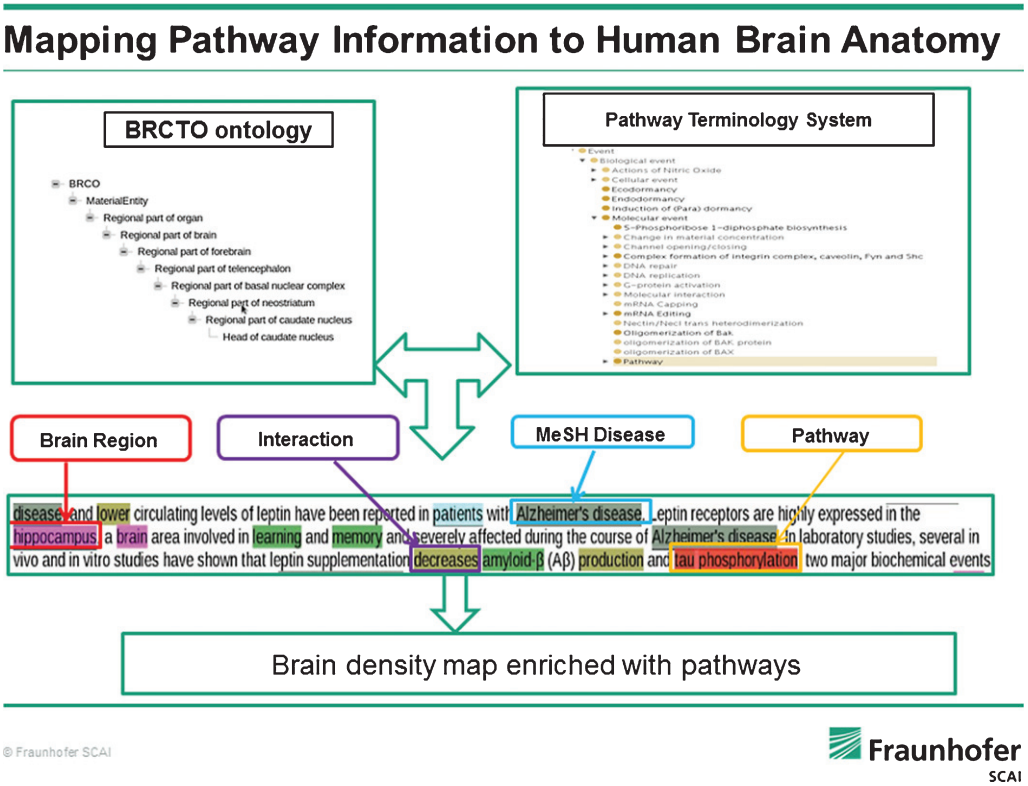


Fig. 3. Brain density map. This figure depicts the tri-occurrence based workflow for capturing pathways and events and mapping them to the anatomy of brain regions.

Disease:"Alzheimer Disease"]) AND [Interaction]) AND [PTS:"Pathway"]. We used the Interaction dictionary for capturing "qualifiers" of pathways such as perturbed, increased or inhibited and manually curated the list of resulting pathway – brain region co-occurrences. Similarly, we used the following query for retrieving articles pertaining to healthy brain: "((([BRCT:"Hippocampus"]) AND [Organism:"Homo sapiens"]) NOT [MeSH Disease]) AND [PTS:"Pathway"]".

Drug mode of action analysis

We performed a tri-occurrence search using PTS, a dictionary file from the HBP database along with the MeSH ontology to find relevant pathways in the literature in the context of AD and PD. Once we identified the top ranking drugs for both disease scenarios, we searched for the mechanism of action of those drugs in DrugBank. Although our initial analysis was focused on AD, we did not find any drugs with unclear mechanism of action. Therefore, we shifted our focus to PD. We identified Rasagiline to have an unclear mechanism of action. Consequently, we searched SCAIView using the query "(([Drug Names: "Rasagiline"]) AND [PTS]) AND [MeSH Disease: "Parkinson Disease"]". Figure 4 provides a schematic representation of a drug and a pathway dictionary, when combined can help in understanding the mode of action of a drug and its effect on the pathways resulting in diseases.

RESULTS

PTS system description and evaluation

The PTS consists of 6,596 classes, organized in a hierarchical structure. Each class has been annotated, whenever available, with additional information such as definition, source identifiers, and synonyms. Figure 5 illustrates the PTS root classes, their organization and exemplary annotations inside the Protégé environment. Figure 6 shows an overview on the PTS class provenance and the contribution of the integrated ontologies and terminologies. The PTS file in OWL format can be freely accessed, browsed

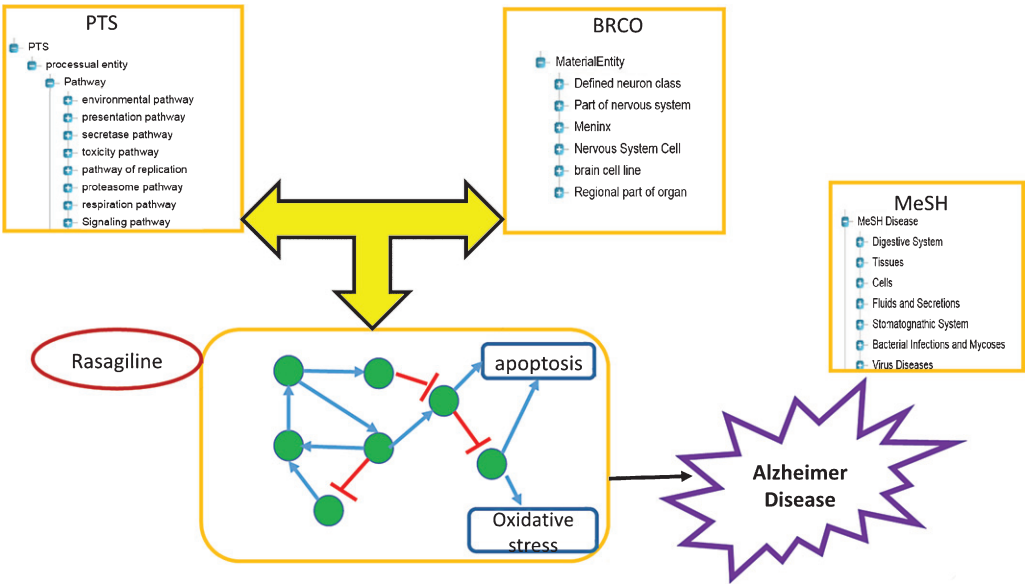


Fig. 4. Finding the mode of action through tri-occurrence approach. This figure is a cartoon representation of the drug, pathway and MeSH disease dictionary. The figure explains how the combination of three dictionaries can help in understanding the effect of the drugs on pathways resulting in different diseases.

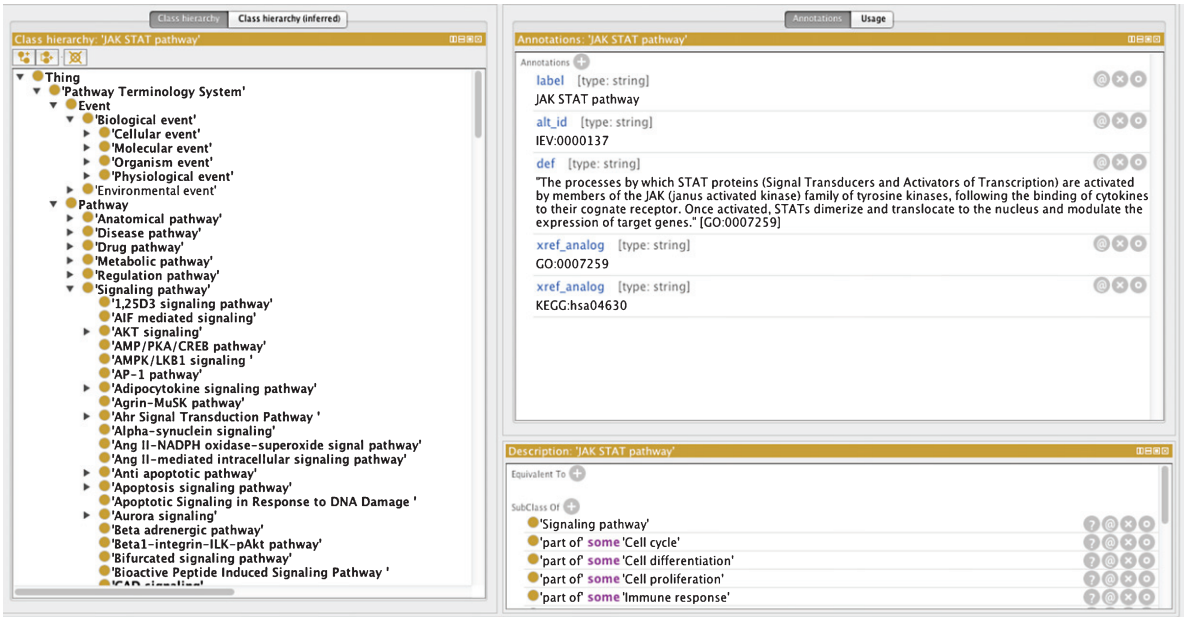


Fig. 5. Representation of PTS in Protégé. This figure is a snapshot of the PTS, displayed in Protégé ontology editor, illustrating hierarchical organization of the PTS in the left and annotation field of the JAK-STAT Pathway, as an example, in the right.

and downloaded through the BioPortal repository at http://bioportal.bioontology.org/ontologies/PTS.

The performance of the PTS system was evaluated as described in the Methods section. For abstracts, the evaluation resulted in a recall of 0.83, a precision of 0.83, and an F-score of 0.83. These values are acceptably high for the system to be used for applications based on PubMed abstracts. For full texts, recall was 0.8, precision 0.63, and F-score 0.71. The overall IAA rate was at a satisfactory level of 81.51% (74.58%

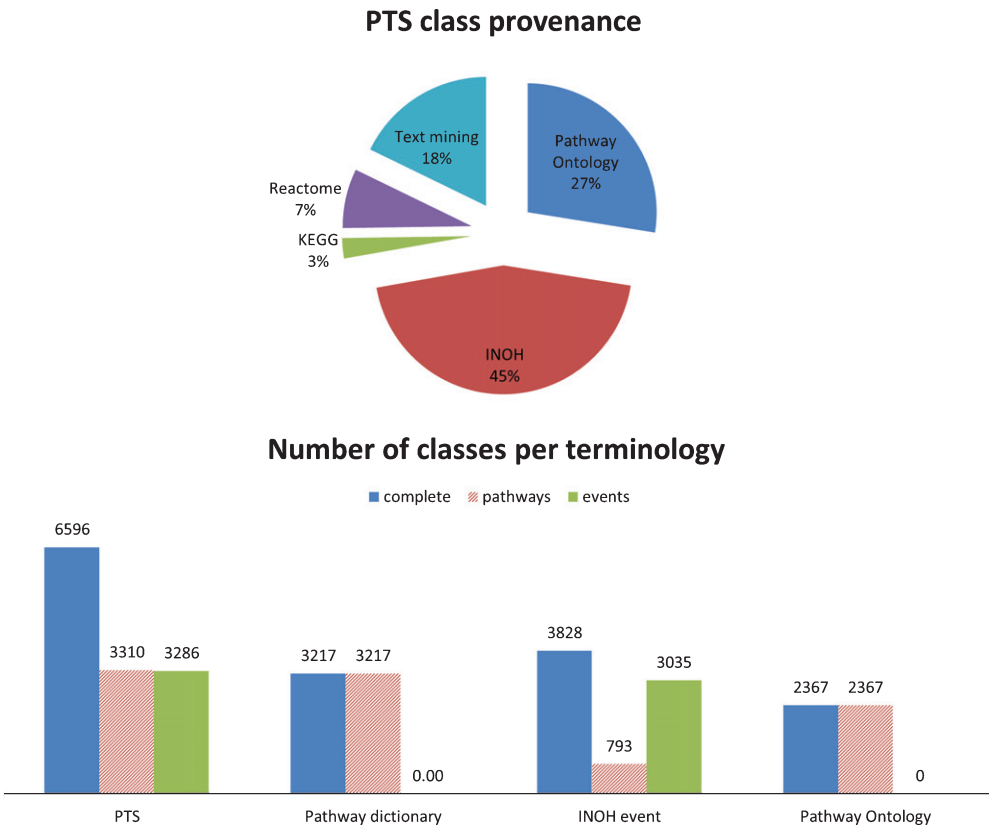


Fig. 6. Top: Provenance of PTS classes. The figure shows the contribution of the integrated ontologies and terminologies to the PTS. 45% of classes come from INOH, 27% from the Pathway Ontology, 18% come from text mining, and 10% from pathway databases. Bottom: Number of classes in the PTS and its contributing ontologies/terminologies, split up into pathway and event classes. After merging and re-organizing classes under their pathway and event parent classes, the PTS consists of 3,310 pathways and 3,286 events. Please note that the Pathway dictionary already includes Pathway Ontology classes, classes from text mining and public databases and is an interim step towards the PTS.

for "Event" classes versus 90.21% for "Pathway" classes). For comparison, our separate evaluation of the Pathway Ontology integrated with pathway mentions from public databases and text mining alone (i.e., without our integration in the PTS and INOH, thus without any events that participate in the pathways) resulted in an F-score of 0.76 (recall: 0.74, precision: 0.77).

Applications of the PTS system

Mapping cellular pathways to the anatomy of human brain

Query of the literature for cellular pathways relevant to the brain anatomy in both healthy and AD conditions, as described in the Methods section, resulted in retrieval of 209 documents for AD and 122 documents for healthy state, respectively (last accessed as of August 18, 2015). Out of the 209 documents, 39 documents contained information about the mode of regulatory pathways. The extracted documents were further manually curated for relevant pathway – brain region co-occurrences for both healthy and disease states. The results are summarized in Tables 1 and 2. Table 1 represents evidence supporting pathway-brain associations in healthy state (6 associations) whereas Table 2 represents similar associations for the disease state plus the type of pathway perturbations under the disease condition (16 associations).

Following the comparative analysis of cellular pathways associated to healthy and diseased brain cells and regions, we tried to visualize the obtained results in association with brain regions. Therefore, we made use of the TVB environment [50]. Figure 7 shows the BRCO brain regions and cell types together

with their mapped PTS pathways uploaded in the TVB tool. A mapping from the TVB brain regions to the BRCO brain regions empowers the integration of the latter into a TVB project. Please note that the hierarchical structure reflects part-of or located-in (not is-a) relationships in many cases. Figure 8 illustrates pathways in the diseased brain with their state of perturbation as indicated by arrows based on the "interaction" column from Table 2.

Mechanistic analysis of pathways targeted by drugs

During drug discovery and development, knowing which disease-related pathways are to be modified by the candidate drug is the key for successful efficacy and safety outcomes. Despite this, it is surprising that the mechanism of action for 18% of FDA-approved drugs is not clear [51]. On one hand, well-established drug databases such as DrugBank do not provide sufficient information about the mechanistic interference of drugs with their targeted pathways, and, on the other hand, information on the mode of action of drugs is often buried under the exponentially growing number of publications. The following proof-of-concept application demonstrates the potential use of the PTS system to extract relevant pieces of evidence from the literature that unravel the mode of action of the drug Rasagiline, which is a top selling drug prescribed for PD and does not have a clear mode of action reported in DrugBank.

The existing information in DrugBank suggests that this drug exerts an inhibitory effect on the MAO-B target protein, which leads to increased dopamine levels and enhanced dopaminergic activity in striatum (Fig. 9). However, the precise molecular mechanism behind this effect is unclear. In order to assemble relevant publications for pathways involved in the mode of action of this drug, we retrieved 81 relevant articles using the query mentioned in the Methods section that report substantial experimental evidence on the mode of action of Rasagiline. As illustrated in Fig. 10, we were able to considerably increase the resolution of pathway information describing the mode-of-action of Rasagiline when compared to Fig. 9. Neuronal death (apoptosis) underlies many neurodegenerative diseases such as PD and AD, which is triggered by a couple of perturbations, particularly in mitochondrial functioning [52]. The monoamine oxidases are known to be involved in the process of apoptotic signaling in mitochondria. According to the improved pathway model in Fig. 10, monoamine oxidase A (MAO-A) induces neural apoptosis by binding to a dopaminergic neurotoxin while monoamine oxidase B (MAO-B) produces hydrogen peroxide by oxidizing dopamine in brain, which results in

Table 1
Overview of brain-specific cells and regions associated with cellular pathways in the healthy human brain. Each association is supported by a reference to PubMed identifier of the published evidence

Brain Region	Pathway	PMID
Hippocampus	PKA signaling	21483429
Hippocampus	Neuronal pathway	24655332
Hippocampus	Long-term potentiation	1350592
Hippocampus CA3 pyramidal cell	Long-term potentiation	19596521
Dentate gyrus	Glutamatergic pathways	21035522
Cornu ammonis	Glutamatergic pathways	21035522

Table 2
Overview of brain-specific cells and regions associated with cellular pathways in the Alzheimer's disease state. Each association is supported by a reference to PubMed identifier of the published evidence

Brain Region	Pathway	Interaction	PMID
Hippocampus_CA_Pyramidal Cell	Mitochondrial Biogenesis Pathway	Disrupted	24448779
Hippocampus	Energy Metabolism	Reduced	9689449
Hippocampal Neurons	Nogo-A to Nogo-66 receptor (NgR)Pathway	Overexpressed	22139298
Superior Temporal Cortex	Glucose Metabolism	Diminished	1719135
Hippocampus	Protein phosphorylation	Associated	21515431
Hippocampus	Leptin Signaling	Decreases	23383396
Hippocampus	Glucose Metabolism	Decreases	23383396
Hippocampus	PI3K/Mtor Pathway	Decreases	23383396
Hippocampus	JAK/STAT Pathway	Decreases	23383396
Hippocampus	ERK Pathway	Decreases	23383396
Hippocampus	p38 MAPK pathway	Significant Increase	11677259
Hippocampal Neurons	APP Processing	Increased neuritic "branching"	10723071
CA3	Tau Phosphorylation	Tau aggregation	21677375
Hippocampus	Oxidative stress pathway	Induces	20863531
Hippocampal neurons	Endocytosis pathway	Downregulation of flotillin-2	18337418
Hippocampal formation	Glutamatergic pathway	Increase in GluR1 and GluR2/3	8773259

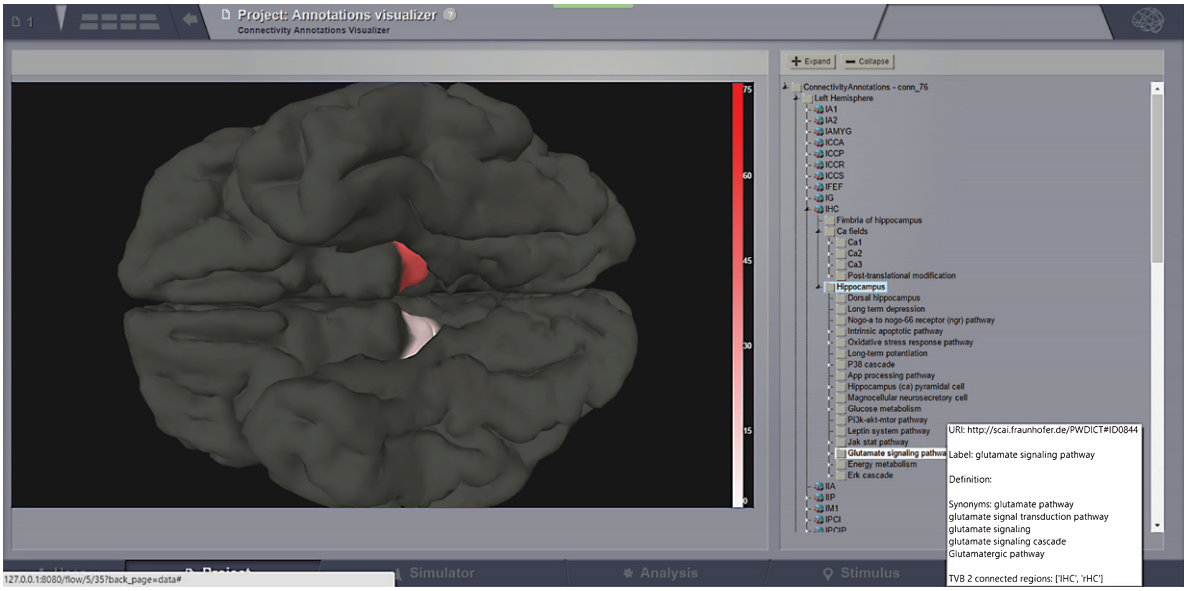


Fig. 7. Visualization of the BRCO brain regions and cells together with the PTS pathways in the TVB's Brain Activity Visualizer. Note that the hierarchy of terms reflects both is-a and located-in or part-of relationships.

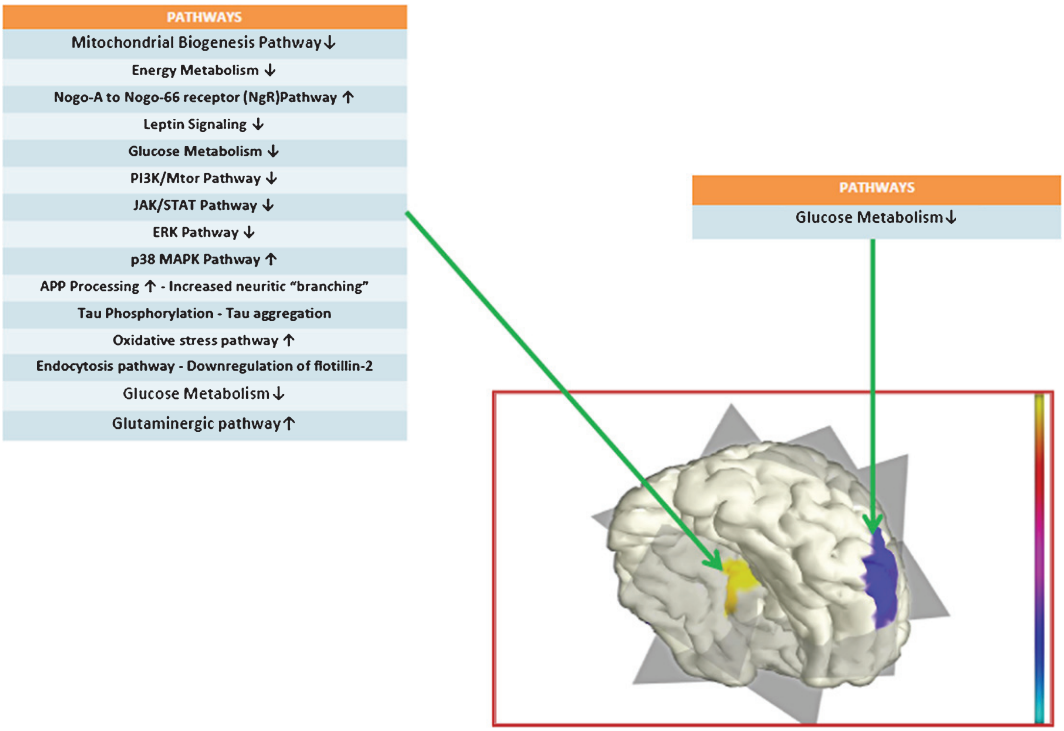


Fig. 8. Cellular pathways active in AD brain, visualized using TVB's Brain Activity Visualizer. Visualization of altered cellular pathways in the affected hippocampal (yellow) and superior temporal (blue) cortices in AD brain. Arrows denote upregulation (↑) or downregulation (↓) of pathways under disease condition in these brain regions.

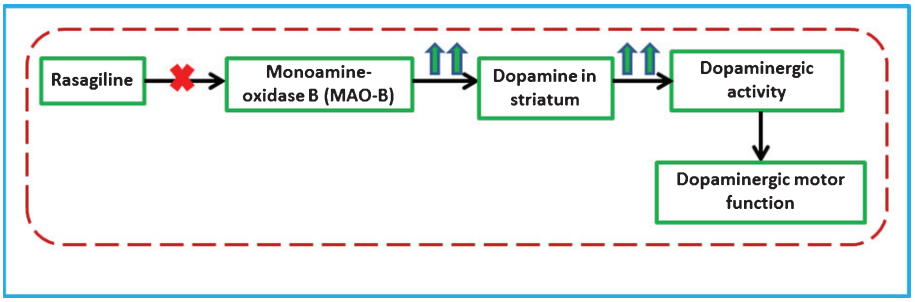


Fig. 9. Representation of mode of action of Rasagiline from DrugBank. A cartoon version of the description provided in the DrugBank for the mode of action of Rasagiline, a top-selling Parkinsonian drug.

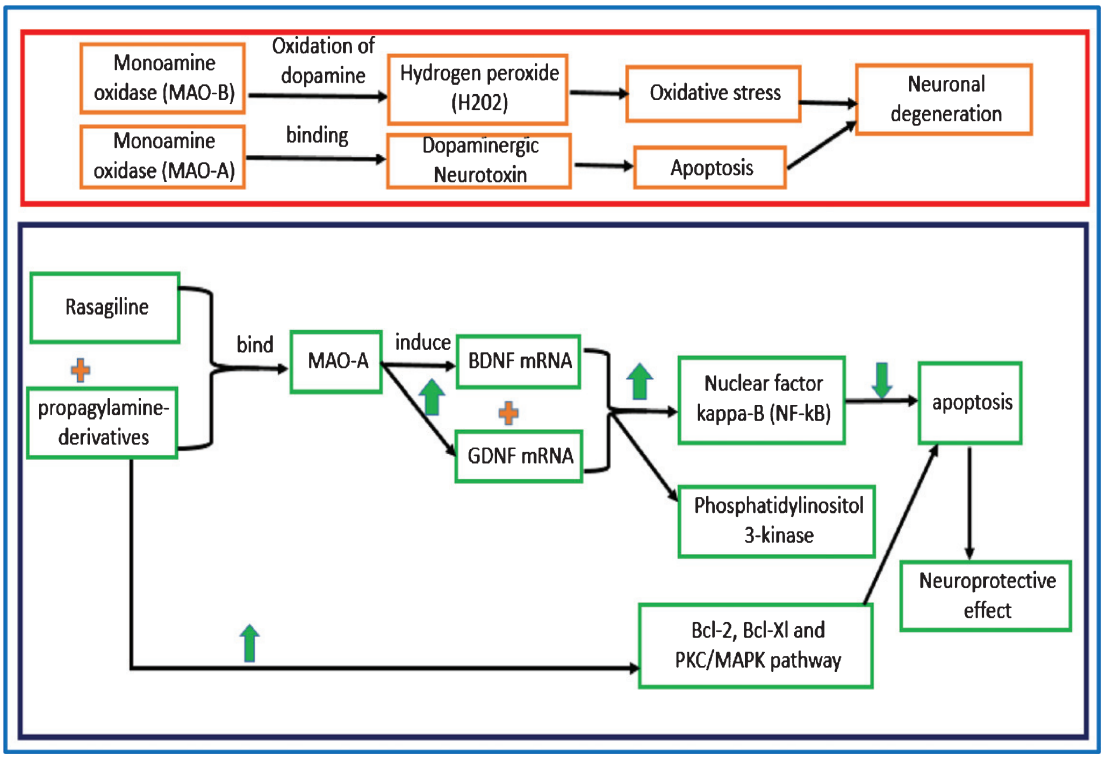


Fig. 10. Representation of mechanistic inference of mode of action of Rasagiline with neurodegenerative processes. The processes colored in red summarize the mechanistic role of MAO genes in neurodegeneration whereas those in green explain how Rasagiline mechanistically interferes with neurodegenerative processes in PD.

oxidative damage to neurons in PD. Rasagiline and other MAO-B inhibitors prevent the apoptotic cascade by inducing genes such as bcl-2 and glial cell derived neurotrophic factor [53]. The Rasagiline effect is primarily mediated by the suppression of the cell death cascade initiated by pro-apoptotic mitochondrial proteins. Furthermore, the binding of Rasagiline with propargylamine derivatives protects mitochondrial viability. This association causes the activation of neurotrophic factors such as Brain derived neurotrophic factor (BDNF) and Growth derived neurotrophic factor (GDNF), which further activate nuclear factor kappa B (NF-κB), P3K and PKC/MAPK pathways, leading to reduced apoptosis and increased neuroprotective effects [54].

Discovery of pathway knowledge by fingerprinting patents

Patents are major sources of novel research findings and they contain rich information on new data and innovative technologies so that they have been previously exploited for analysis of technological change impact on informatics advances [55], extraction of chemical interactions [56], and identification of competitors information for drug discovery and development [57]. Here, we show that the PTS lexicon in combination with our fingerprinting method can be successfully applied to mining patents for cellular pathways that have been reported in the context of AD and their corresponding genes and drugs.

Using the similarity matrix described in the Methods section, we selected those patents with high dissimilarity scores (i.e., patents that contain pathway information other than the two well-known pathways "APP processing" and "Amyloidogenic pathway") so that we could identify potentially novel, less known pathways that are linked to a gene or drug action in the context of AD. As a result, we were able to retrieve 22 pathway-gene/drug co-mentions from three patents, respectively. Out of the 22 pathway-gene co-mentions, there were 6 pathways with gene and drug co-mentions, namely malate-aspartate shuttle, Glucuronidation, Ubiquitin mediated proteolysis, Homeostasis, Calcium signaling pathway, and Acetylcholine signaling pathway (Table 3). These pathways have been less investigated in the context of AD; for example, we found only one publication (PMID: 25766789 published in 2015) reporting on a potential role for malate-aspartate shuttle pathway in AD. Similarly, for the role of NGF in acetylcholine signaling pathway under AD conditions, we found only one publication that explicitly relates NGF activity to acetylcholine (PMIDs: 24948063).

Qualitative pathway analysis versus quantitative pathway enrichment

Current pathway analysis methods, including GSEA, provide an overview of pathways for which differentially expressed genes are enriched. However, the biological interpretation of expression data does often not go beyond pathway identification at an abstract level so that further details on how these individual genes contribute to tissue-specific perturbed signaling under disease condition remains unclear. In this scenario, our aim is to demonstrate how a knowledge-based pathway analysis approach compared to the popular statistically driven GSEA method brings more biological context and information resolution to mechanistic interpretation of biological data. We used the work of [58] who analyzed differentially expressed genes in various brain regions of AD subjects. Accordingly, the authors report the following pathways and processes as top hits for expressed genes in entorhinal cortex (EC) and hippocampus (HIP): Cellular physiological process (EC & HIP), transport (EC), cellular process (EC), synaptic transmission (EC), intracellular transport (HIP), establishment of cellular localization (HIP), cellular localization (HIP), and metabolism (HIP). We picked the top 30 genes from their list of statistically significant genes for EC and HIP regions, and performed both, GSEA analysis and PTS-based literature analysis. GSEA results indicated that differentially expressed genes in EC were statistically enriched for the Neuroactive ligand-receptor interaction pathway with only 4 genes in overlap and in HIP for the Neurotransmitter release cycle pathway with 5 genes in overlap. In contrast, our PTS-based search in the literature led to identification of specific pathways for 9 genes from the EC list and 8 genes from the HIP list. These results are summarized in Table 4 and supported by

Table 3
Unique pathway-gene/drug co-mentions from patent fingerprinting

Patient ID	Pathway	Gene	Drug
US20100279943	Malate_aspartate shuttle	MDH1	
US20100279943	Malate_aspartate shuttle	SOD1	
US20110130392	Glucuronidation		N-acetylcysteine
US20110130392	Glucuronidation	UGT1A7	
US20110130392	Glucuronidation	UGT1A7	
US20110130392	Glucuronidation	UGT1A10	
WO2013004527A1	Ubiquitin-mediated proteolysis	IDE	
WO2009133142A1	Homeostasis	ATP2B1	
WO2009133142A1	Calcium signaling pathway	APP	
WO2009012571A1	Acetylcholine signaling pathway	NGF	

Table 4
Gene expression analysis using PTS

Brain Region	Differentially Expressed Genes	PTS Results
EC	ITPKB (upregulated)	Neuronal cell apoptosis, APP processing pathway (PMID: 24401760); Calcium signaling pathway (PMID: 16257224)
EC	GRIN1 (downregulated)	Long-term potentiation (PMID: 24858312)
EC	ALDOA (downregulated)	Glucose metabolism (PMID: 20866111)
EC	G6PD (downregulated)	Pentose phosphate pathway (PMIDs: 10510282; 10392540)
EC	CHRM1 (downregulated)	APP processing pathway (PMID: 20335454; 11193170; 19906975)
EC	EEF1A2 (downregulated)	Glucose metabolism (PMID: 16782585)
EC	NTSR2 (downregulated)	Energy homeostasis (PMID: 20858966)
EC	MIF (downregulated)	Extracellular signal-regulated kinase-1/2 MAP kinases pathway (PMID: 18786268)
EC	SRSF3 (downregulated)	Cholesterol homeostasis (PMID: 20232416)
HIP	THRA (upregulated)	Amyloidogenic pathway (PMID: 20403092)
HIP	MAGI2 (upregulated)	Protein degradation pathway (PMID: 19668339)
HIP	HRK (upregulated)	Neuronal cell death (PMID: 24420784, 16524368)
HIP	MAP1B (upregulated)	Neuronal death pathway (PMID: 16234245, 12376528)
HIP	RTN3 (downregulated)	APP processing pathway (PMID: 17032350)
HIP	FABP3 (downregulated)	Fatty acid metabolism and lipid transport (PMID: 24088526)
HIP	SNAP25 (downregulated)	Neurotransmitter secretion (PMID: 11923424, 18194215, 11515747)
HIP	HSPA8 (downregulated)	Chaperone_mediated autophagy pathway (PMID: 20697033, 22277499)

publication identifiers as references to the corresponding evidence.

These results indicate that the PTS-based pathway analysis provides context-sensitive information on involvement of differentially expressed genes in various signaling pathways underlying the disease mechanism.

DISCUSSION

Physiological pathways act as key players in understanding the etiology of disease. Although several databases host information of such pathways, their content is either devoid of biological context (e.g., cell type- or tissue-specific activity, disease-specific alterations) or biased toward a particular disease condition (e.g., cancer). Consequently, when investigating pathways perturbed under brain disorders, particularly neurodegenerative diseases, researchers often face difficulties interpreting their data in the right context. Thus, biomedical literature is an alternative resource to obtain context-sensitive information of such pathways and their components. But finding key pathway mechanisms in the literature is often challenging due to the unstructured nature of textual information. Moreover, many pathways are not explicitly mentioned as 'pathway' in the literature but as biological processes or chains of events. This prompted us to address this challenge by adding cellular events to the PTS, which provides a powerful lexicon for identification of these entities in the text. Beside various types of pathways that have been classified under different categories, biological events such as cellular, molecular and physiological events have been also included to ensure the coverage of all possible cellular or molecular processes that are not explicitly considered as pathways in the literature but represent similar chains of interactions. The functional evaluation of the PTS lexicon indicated that its performance for detecting pathways and events in publications was good enough to be used for information retrieval and extraction. The reasonably high values of F-scores verify that the PTS lexicon has a good coverage of pathway-related terms, which can be captured with a good precision in the text. We are aware that using a combination of dictionaries as we did in our analyses (e.g., the PTS together with the BRCO) lowers the combined F-score. We suppose, however, that the co-occurrence of important entities will usually be found in more than one abstract, so recall will not be affected so much for the important term combinations.

After evaluation, the PTS was applied to mining the literature including biomedical publications and patents. We believe that the first step toward systems analysis of disease data is to set the cellular or tissue context by which such data become biologically meaningful. To the best of our knowledge, there is no reference compendium or inventory of pathways specific to human brain anatomic regions neither for normal nor for disease conditions. The importance of a pathway inventory for the normal brain lies in the fact that any conclusions drawn from a disease model without comparison with a baseline, normal functioning model is inconclusive. Hence, generation of a brain pathway map for both healthy and

diseased states provided us with an opportunity to perform a comparative analysis. We observed that out of several pathways that are active in the healthy state, glutamatergic pathway has been reported to be perturbed in the disease state. Glutamatergic pathways are main excitatory neurotransmitters; particularly NMDA glutamate receptors are reported to be involved in regulating hippocampal plasticity. These receptors are sites of constant transmission of Ca2+ and Mg2+ ions, which help in maintaining the brain plasticity. However, in the disease state, glutamate receptors induce constant activation of the NMDA receptor resulting in continuous influx of Ca2+ ions, and ultimately neuronal injury and death [59].

In the disease state, our results indicate that a set of different pathways are reported to be perturbed, which mainly affect signaling and energetics at hippocampus. For instance, the leptin signaling pathway has been found to play a role in amyloid-β accumulation in AD. In fact, the leptin signaling pathway functions by binding the leptin receptor and this binding modulates other pathways such as JAK/STAT, ERK, and PI3K/Akt/mTOR pathways, thereby downregulating amyloid-β production and tau phosphorylation [60]. It has been shown that high levels of leptin in blood is associated with a lower risk of AD, which is in line with the downregulation of leptin signaling pathway in the hippocampus of AD brains. Moreover, [61] have identified leptin signaling pathway at the core of 7 converging hormone-signaling pathways that are perturbed under dementia and functionally validated its role. Accordingly, leptin signaling pathway appears to be a candidate target for interfering with the disease progression.

In contrast to the above scenario, the process of drug discovery often starts with screening for lead compounds and thereby mechanisms behind target candidates remain unknown. Thus, prior knowledge about mechanistic action of the candidate drug and its interference with possible molecular pathways that regulate disease-related processes and outcomes becomes highly valuable. Collecting, curating, and modeling this knowledge helps to understand drug mode-of-action and to support existing hypotheses on the mode-of-action with a high-resolution causal model. As we demonstrated in the case of Rasagiline, the poor annotation of mode-of-action for drugs in public databases such as DrugBank can be significantly improved when contextual information is added to the current understanding of mechanism of action. This proof-of-concept demonstration implies that the wealth of hidden knowledge in the literature can be systematically collected and modeled using the PTS lexicon to shed light on mechanistic aspects of targeted pathways and biological processes.

Indeed, the value of systematic collection and modeling of the pathway knowledge in the literature can go beyond re-construction of known pathways from the scientific publications so that the aim will be to perform knowledge discovery in patents for those disease pathways that are less known to the public. The idea behind this notion is that novel findings reported in patents may serve as sources of information for filling the gaps in our understanding about disease mechanism. As demonstrated by application of the PTS to patents relevant to AD, there is a chance to find pathways and mechanistic information in patents that can provide new insights into the underlying pathophysiology. For instance, in our study it was interesting to note that apart from acetylcholine signaling pathway and ubiquitin mediated proteolysis, other pathways such as malate aspartate shuttle and glucouronidation pathway are not widely reported in association with AD. In fact, malate dehydrogenase (MDH) plays a key role in the malate/aspartate shuttle by catalyzing the interconversion of malate and oxaloacetate in cytoplasm and mitochondria [62]. In case of AD, the catalytic enzyme MDH forms complex with protein mutants such as SOD1 and MDH1, which are associated with neurodegenerative diseases. When we tried to support our findings with published data, we only found one recently published article that indirectly addressed the role of the malate aspartate shuttle in AD [63]. These results suggest that the PTS-based fingerprinting of patents in a systematic manner provides an opportunity to find new directions in the AD research and get away from heavily researched genes and pathways. We have previously shown that, on the basis of chronological analysis of literature, AD research is biased to a handful of genes and pathways such as MAPT and the APP processing pathway [64].

Currently a major limitation in neurodegeneration research is the lack of an inventory or compendium of well-curated, context-sensitive, and disease-specific signaling pathway models that can be used for the interpretation of molecular data such as gene expression data. In the absence of such an inventory, there is always the danger of data misinterpretation and bias towards predominant cancer mechanisms when using existing pathway information in public repositories. Moreover, our knowledge about pathways in the

brain is limited and a systematic comparison of gene expression data combined with GSEA and the pathway inventory—as demonstrated earlier at a smaller scale—should give us a clue on how much information we have or lack about signaling pathways in the human brain. Indeed, the caveat of statistical methods such as GSEA is identification of pathways in a very abstract level from a combination of deregulated genes, which may under-represent mechanistic signatures (i.e., disease-specific processes); for instance, we have been able to detect under-represented signatures of hormone signaling in pathways enriched for dementia-related proteins. In contrast, annotation of pathways to individual expressed genes not only clearly links genes to their specific role in the disease context (e.g., in hippocampus the gene HRK is involved in the neuronal cell death pathway) but also correlates expression levels of those genes to their corresponding pathways (e.g., HRK is upregulated in hippocampus and triggers the neuronal cell death pathway).

The advantage of having such an inventory is manifold: It represents our current knowledge about known signaling pathways in various regions of the brain; it can serve as a baseline for connecting dots (i.e., unknown parts) between signaling pathways using signals from experimental data such as gene expression data in a certain brain region or cell type for which no prior knowledge exists in the literature; it provides curators of pathway information with the possibility to quickly find detailed information in the literature for construction of pathway cartoons and their deposition into pathway databases; and finally, it prepares the ground for mechanistic, cause-and-effect representation of pathways that go beyond static, cartoon-like representation. Presently, signaling pathways and cellular events are mainly represented in cartoons and drawings so that they are not much interoperable and amenable to computational analysis (e.g., reasoning and simulation). As the launch of several initiatives such as the Human Connectome Project [65] in the US and AETIONOMY as well as EMIF-AD [66] projects in EU indicate, research on brain and neurodegenerative diseases is moving toward integrative analysis of structural, functional, and clinical aspects of the brain system and its disorders. To this end, we see a need for a compendium that maps signaling pathways to their underlying anatomical structure of brain. Such a reference map can have important implications for bridging the gap between clinical outcomes (e.g., measured by imaging technologies) and molecular data (e.g., OMICs data). We have undertaken the effort to use the pathway mining strategy described in this work for retrieving and extracting signaling information from biomedical text and consolidating this information for possible future use in cause-and-effect network models that are computer-processable. Moreover, we plan to build a dedicated database for depositing these pathway models and propose a new paradigm in enrichment analysis of gene sets or other sets of biological entities using these causal models.

ACKNOWLEDGMENTS

We thank Alexander Klenner (formerly Fraunhofer SCAI, now European Patent Office, The Netherlands) for his fruitful contributions and ideas to the UIMA patent fingerprinting pipeline. We also thank Shweta Bagewadi (Fraunhofer SCAI) for providing us with the patent corpus used for fingerprinting and her explanations on how the corpus was generated. The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n°115568 (project AETIONOMY), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. Further support was provided by the James S. McDonnell Foundation (Brain Network Recovery Group JSMF22002082).

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/15-1178r1).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-151178.

REFERENCES

[1] Yuryev A (2008) *Pathway Analysis for Drug Discovery: Computational Infrastructure and Applications*. John Wiley & Sons, Hoboken, NJ.

[2] Kyoto Encylopedia of Genes and Genomes, http://www.genome.jp/kegg/, Accessed on December 9, 2015.

[3] BioCarta, http://www.genecarta.com/, Accessed on December 9, 2015.

[4] WikiPathways, http://www.wikipathways.org/index.php/WikiPathways, Accessed on December 9, 2015.

[5] Adriaens ME, Jaillard M, Waagmeester A, Coort SL, Pico AR, Evelo CT (2008) The public road to high-quality

curated biological pathways. *Drug Discov Today* **13**, 856-862.

[6] Khatri P, Sirota M, Butte AJ (2012) Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Comput Biol* **8**, e1002375.

[7] Pathway Ontology, https://bioportal.bioontology.org/ontologies/PW, Last updated on October 27, 2015.

[8] Rat Genome Database, http://rgd.mcw.edu/, Last accessed on December 9, 2015.

[9] Petri V, Jayaraman P, Tutaj M, Hayman G, Smith JR, De Pons J, Laulederkind SJ, Lowry TF, Nigam R, Wang S-J, Shimoyama M, Dwinell MR, Munzenmaier DH, Worthey EA, Jacob HJ (2014) The pathway ontology – updates and applications. *J Biomed Semantics* **5**, 7.

[10] 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.

[11] Neuroscience Information Framework (NIF) Standard Ontology,https://bioportal.bioontology.org/ontologies/NIFSTD, Last updated on October 9, 2015.

[12] The Virtual Brain, http://www.thevirtualbrain.org/, Last accessed on December 9, 2015.

[13] Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH (2013) Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med* **5**, 48.

[14] Hodges A (2006) Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet* **15**, 965-977.

[15] INOH Ontology, http://www.berkeleybop.org/ontologies/iev.owl, Last accessed on December 9, 2015.

[16] Grenon P, Smith B (2004) SNAP and SPAN: Towards dynamic spatial ontology. *Spat Cogn Comput* **1**, 69-103.

[17] Tenenbaum D (2013) KEGGREST: Client-side REST access to KEGG. R package version, 1(1).

[18] Reactome Database, http://www.reactome.org/, Last accessed on December 9, 2015.

[19] Pathway Interaction Database, http://pid.nci.nih.gov/, Last accessed on December 9, 2015.

[20] Hanisch D, Fundel K, Mevissen H-T, Zimmer R, Fluck J (2005) ProMiner: Rule-based protein and gene entity recognition. *BMC Bioinformatics* **6**(Suppl 1), S14.

[21] Lipscomb CE (2000) Medical subject headings (MeSH). *Bull Med Libr Assoc* **88**, 265-266.

[22] Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS (2011) DrugBank 3.0: A comprehensive resource for 'Omics' research on drugs. *Nucleic Acids Res* **39**, D1035-D1041.

[23] Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, Rosenstein MC, Wiegers TC, Mattingly CJ (2013) The Comparative Toxicogenomics Database: Update 2013. *Nucleic Acids Res* **41**, D1104-D1114.

[24] Zhu F, Shi Z, Qin C, Tao L, Liu X, Xu F, Zhang L, Song Y, Liu X, Zhang J, Han B, Zhang P, Chen Y (2012) Therapeutic target database update 2012: A resource for facilitating target-oriented drug discovery. *Nucleic Acids Res* **40**, 1-9.

[25] Kuhn M, Szklarczyk D, Franceschini A, von Mering C, Jensen LJ, Bork P (2012) STITCH 3: Zooming in on protein-chemical interactions. *Nucleic Acids Res* **40**, D876-D880.

[26] Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P (2010) A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* **6**, 1-6.

[27] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res* **40**, D1100-D1107.

[28] Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* **35**, D198-D201.

[29] Roth BL, Lopez E, Patel S, Kroeze WK (2000) The multiplicity of serotonin receptors: Uselessly diverse molecules or an embarrassment of riches? *Neuroscientist* **6**, 252-262.

[30] Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucl Acids Res* **28**, 27-30.

[31] Groth P, Pavlova N, Kalev I, Tonov S, Georgiev G, Pohlenz H-D, Weiss B (2007) PhenomicDB: A new cross-species genotype/phenotype resource. *Nucleic Acids Res* **35**, D696-D699.

[32] European Prevention of Alzheimer's Dementia Consortium (EPAD), http://www.imi.europa.eu/content/epad, Last accessed on December 9, 2015.

[33] Aetionomy, http://www.aetionomy.org/, Last accessed on December 9, 2015.

[34] Gardner D, Bowden DM (2009) NIH Public Access. *Neuroinformatics* **6**, 149-160.

[35] Bowden DM, Song E, Kosheleva J, Dubach MF (2012) NeuroNames: An ontology for the BrainInfo portal to neuroscience on the web. *Neuroinformatics* **10**, 97-114.

[36] Gremse M, Chang A, Schomburg I, Grote A, Scheer M, Ebeling C, Schomburg D (2011) The BRENDA Tissue Ontology (BTO): The first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res* **39**, D507-D513.

[37] Bodenreider O (2004) The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Res* **32**, D267-D270.

[38] Tsarkov D, Horrocks I (2006) FaCT++ Description Logic Reasoner: System description. In *Proceedings of the Third International Joint Conference (IJCAR 2006)*, pp. 292-297.

[39] Brain Region and Cell Type Terminology (BRCT), http://bioportal.bioontology.org/ontologies/BRCT, Last updated on November 3, 2014.

[40] Hofmann-Apitius M, Fluck J, Furlong L, Fornes O, Kolárik C, Hanser S, Boeker M, Schulz S, Sanz F, Klinger R, Mevissen T, Gattermayer T, Oliva B, Friedrich CM (2008) Knowledge environments representing molecular entities for the virtual physiological human. *Philos Trans A Math Phys Eng Sci* **366**, 3091-3110.

[41] SCAIView Academia, http://www.scaiview.com/scaiview-academia.html, Last accessed on December 9, 2015.

[42] Knowtator, http://knowtator.sourceforge.net/, Last accessed on December 9, 2015.

[43] Ogren PV (2006) Knowtator: A Protégé plug-in for annotated corpus construction. *Proc 2006 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol companion Vol. Demonstr.*, pp. 273-275.

[44] Wu Z, Palmer M (1994) Verbs semantics and lexical selection. *ACL '94 Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pp. 133-138.

[45] PatBase, https://www.patbase.com/login.asp, Last updated on December 9, 2015.
[46] ChemoCR, http://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/products/chemocr.html, Last accessed on December 9, 2015.
[47] PDFBox, https://pdfbox.apache.org/, Last accessed on December 9, 2015.
[48] Ogren PV, Bethard SJ (2009) Building test suites for UIMA components. *Proc Work Softw Eng Testing Qual Assur Nat Lang Process (SETQA-NLP 2009)* Proceeding, pp. 1-4.
[49] Count-Text Tool, http://textmechanic.com/Count-Text.html, Last accessed on December 9, 2015.
[50] Sanz Leon P, Knock SA, Woodman MM, Domide L, Mersmann J, McIntosh AR, Jirsa V (2013) The Virtual Brain: A simulator of primate brain network dynamics. *Front Neuroinform* **7**, 10.
[51] Gregori-Puigjane E, Setola V, Hert J, Crews BA, Irwin JJ, Lounkine E, Marnett L, Roth BL, Shoichet BK (2012) Identifying mechanism-of-action targets for drugs and probes. *Proc Natl Acad Sci U S A* **109**, 11178-11183.
[52] Ghavami S, Shojaei S, Yeganeh B, Ande SR, Jangamreddy JR, Mehrpour M, Christoffersson J, Chaabane W, Moghadam AR, Kashani HH, Hashemi M, Owji AA, Łos MJ (2014) Autophagy and apoptosis dysfunction in neurodegenerative disorders. *Prog Neurobiol* **112**, 24-49.
[53] Naoi M, Maruyama W, Akao Y, Yi H, Yamaoka Y (2006) Involvement of type A monoamine oxidase in neurodegeneration: Regulation of mitochondrial signaling leading to cell death or neuroprotection. In *Oxidative Stress and Neuroprotection*, Editors: Parvez SH, Riederer P, eds. Springer, Vienna, pp. 67-77.
[54] Weinreb O, Amit T, Bar-Am O, Youdim MB (2007) Induction of neurotrophic factors GDNF and BDNF associated with the mechanism of neurorescue action of rasagiline and ladostigil. *Ann N Y Acad Sci* **1122**, 155-168.
[55] Choi C, Kim S, Park Y (2007) A patent-based cross impact analysis for quantitative estimation of technological impact: The case of information and communication technology. *Technol Forecast Soc Change* **74**, 1296-1314.
[56] Jessop DM, Adams SE, Murray-Rust P (2011) Mining chemical information from open patents. *J Cheminform* **3**, 40.
[57] Grandjean N, Charpiot B, Pena CA, Peitsch MC (2005) Competitive intelligence and patent analysis in drug discovery. Mining the competitive knowledge bases and patents. Drug Discov. *Today Technol* **2**, 211-215.
[58] Liang WS, Reiman EM, Valla J, Dunckley T, Beach TG, Grover A, Niedzielko TL, Schneider LE, Mastroeni D, Caselli R, Kukull W, Morris JC, Hulette CM, Schmechel D, Rogers J, Stephan DA (2008) Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc Natl Acad Sci U S A* **105**, 4441-4446.
[59] Campos-Peña V, Meraz-Ríos MA (2014) Alzheimer disease: The role of Aβ in the glutamatergic system. In *Neurochemistry*, Heinbockel T, ed. ISBN: 978-953-51-1237-2, InTech, doi: 10.5772/57367
[60] Marwarha G, Ghribi O (2012) Leptin signaling and Alzheimer's disease. *Am J Neurodegener Dis* **1**, 245-265.
[61] Younesi E, Hofmann-Apitius M (2013) A network model of genomic hormone interactions underlying dementia and its translational validation through serendipitous off-target effect. *J Transl Med* **11**, 177.
[62] Korolainen M, Goldsteins G, Nyman T, Alafuzoff I, Koistinaho J, Pirttil T (2006) Oxidative modification of proteins in the frontal cortex of Alzheimer's disease brain. *Neurobiol Aging* **27**, 42-53.
[63] Zaroff S, Leone P, Markov V, Francis JS (2015) Transcriptional regulation of N-acetylaspartate metabolism in the 5xFAD model of Alzheimer's disease: Evidence for neuron-glia communication during energetic crisis. *Mol Cell Neurosci* **65**, 143-152.
[64] Malhotra A, Younesi E, Gurulingappa H, Hofmann-Apitius M (2013) HypothesisFinder: A strategy for the detection of speculative statements in scientific text. *PLoS Comput Biol* **9**, e1003117.
[65] Human Connectome Project, http://www.humanconnectomeproject.org/, Last accessed on December 9, 2015.
[66] European Medical Information Framework (EMIF-AD), http://www.emif.eu/emif-ad/about-emif-alzheimers-disease, Last accessed on December 9, 2015.

SUMMARY

Biological knowledge is fundamentally complicated, and over time, a high volume of correspondingly complex and highly heterogeneous data has been generated using high-throughput technologies. The use of ontologies and terminologies is prevalent to a greater extent in the life sciences domain as these can be used for collecting, organizing and storing the vast volume of biological data in a standardized manner. The advantages of using biomedical ontologies are: (i) it enables easy interoperability of resources between linked databases, (ii) an efficient search and query of different resources are possible, (iii) it allows for context-specific information retrieval, and (iv) automatic reasoning of data can be performed.

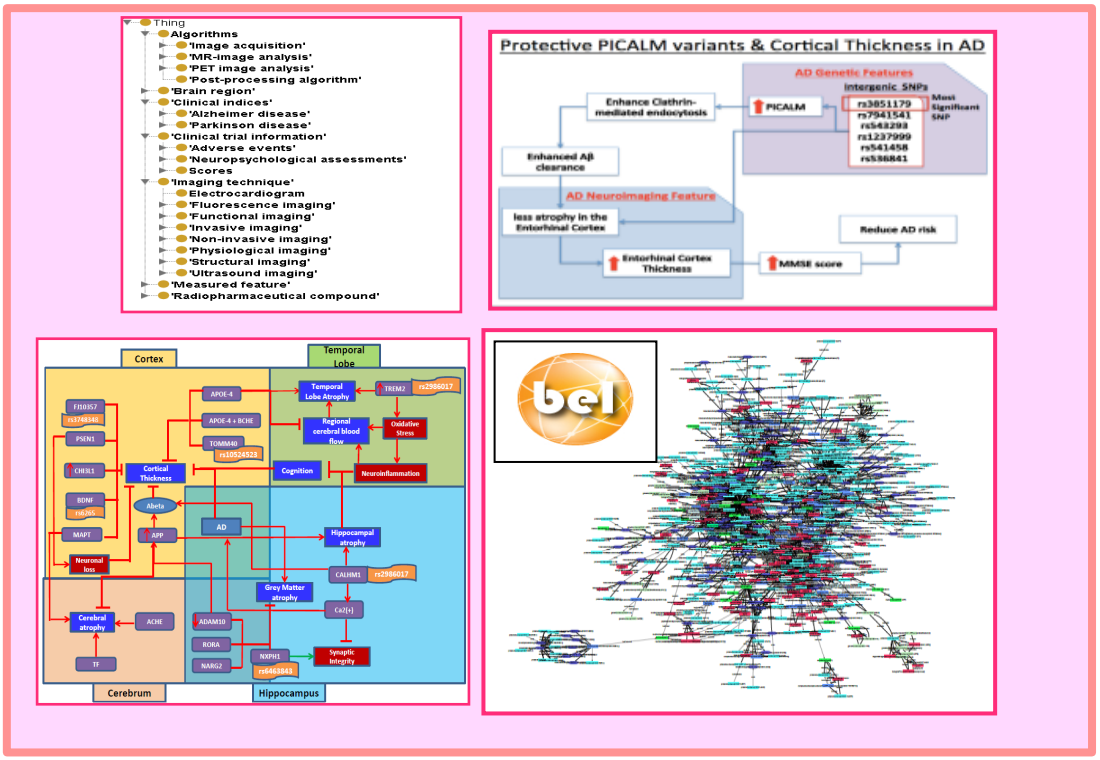
However, with the growing number of molecular databases, one major challenge for the ontology community is in the lack of semantic mapping from a database to an ontology. For instance, there are more than 170 pathway databases which vary widely in form and content; with the multiplicity of information stored in these databases concordant with a lack of quality check over them, raising simple questions often becomes a daunting task for researchers. Furthermore, there also exists a lack of order in the existing pathway ontologies, which motivated us to develop the pathway terminology system (PTS), which combines signaling pathways and biological events to ensure broad coverage of the entire pathway knowledge domain. We have also demonstrated the usability of this system to answer complex questions in any context, especially in the field of NDD.

Molecular pathways consist of the interactions of bimolecular entities triggering a flow of chemical chain-like reactions for regulating or disregarding a biological process. Hence, pathways become an essential target for monitoring disease

progression and for optimized drug treatment. Moreover, they form an integral part of complex, multi-scale modeling approaches, mainly due to the full range of functionality of specific chemical reactions where the effect of genetic and proteomic alterations can forcibly alter metabolic and biochemical reactions, and trespassing of the blood-brain barrier is rendered possible. Additionally, incorporation of context-specific pathway models is highly relevant for gaining more profound insight into bridging the molecular underpinnings of biological processes with significant impact on clinical modalities of the brain.

CHAPTER 3

Mechanistic interpretation of clinical biomarkers in AD



INTRODUCTION

The multimodality of AD is complex and time-dependent, and therefore it has been nearly impossible to accurately diagnose the occurrence of AD with the present genetic, molecular and cellular biomarkers. This is mainly because these biomarkers often are not indicators of the actual state of the disease, but instead, they signify a pathological process (which does not necessarily pinpoint to the actual state of disease progression).

The significance of clinical biomarkers for accurate decision-making in clinical practice has gained considerable momentum in recent years mainly due to global big data initiatives such as ADNI and AIBL. Until recently, clinical readouts and behavioral assessments were confined to clinics and radiological labs for ethical reasons and lack of a reliable platform for enabling data sharing. Such landmark initiatives are primarily launched to address the longstanding problem of linking dynamic clinical outcomes of individual patients to static molecular observations to foster an enhanced understanding of the complex pathology of AD.

This section focuses entirely on exploring the importance of imaging readouts and other clinical biomarkers, bringing order into clinical imaging readouts and the design of a state-of-the-art knowledge model. By integrating clinical biomarkers in a multi-scale computable model, understanding disease mechanisms at a molecular level can allow for accurate diagnosis and treatment of AD.

CONTRIBUTIONS FOR THE MANUSCRIPT

I am entirely responsible for performing all the analysis, data interpretation, and writing of the manuscript: “Complexity across scales: a walkthrough to linking neuro- imaging readouts to molecular processes”.

Sepehr Golrizkhatami contributed Figure 5 of the manuscript.

Journal of Systems and Integrative Neuroscience

Review Article ISSN: 2059-9781

Complexity across scales: a walkthrough to linking neuroimaging readouts to molecular processes

Anandhi Iyappan[1,2], Sepehr Golriz Khatami[1,2], Martin Hofmann-Apitius[1,2*]

[1]Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin 53754, Germany

[2]Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn-Aachen International Center for IT, Dahlmannstrasse 2, 53113 Bonn, Germany

Introduction

Neurodegenerative disorders are often classified as a "multifactorial syndrome" as they share similarities between many genetic, clinical, psychological as well as environmental factors [1,2]. They are highly debilitating clinical conditions that result in progressive neuronal degeneration. Alzheimer Disease (AD) in particular is characterized by progressive neuronal dysfunction and regular decline in cognition and behavior. The cause of AD is broadly classified into two categories: sporadic and familial. The most common form of familial mutations is due to three major genes namely APP, PSEN1 and PSEN2 [3-5]. However, the sporadic form of AD is a complex amalgamation of genetic polymorphisms, environment as well as social lifestyle [6-8]. Although it has been decades since the search for novel biomarkers commenced, there is still no proper diagnosis and treatment for AD [9-11]. Barrett and Hunter's team report that the lack of efficient treatment for AD could be primarily due to a sort of careless misdiagnosis of the disease by physicians [12,13]. Such errors could be an act of lack of attention in routine medical examinations. The existing health care treatment for AD is symptomatic relief [14,15]. However, it is widely disputed that the altering neurodegenerative patterns actually commence much earlier than the actual clinical manifestation of the disease. Therefore, early detection would not only improve the diagnostic accuracy in the clinics but also aid clinicians to offer better and earlier treatment for cognitive and behavioral problems [16,17] as well as better quality of life and economic outcomes.

State-of-the-art brain imaging technologies provide high-resolution information of structural and functional alterations. Therefore, they offer unprecedented early diagnosis; they also provide the opportunity for regular monitoring of a progressive clinical condition such as AD. Furthermore, imaging techniques aid in tracing the transition between diagnostic states such as Mild Cognitive Impairment (MCI) and AD.

Depending on brain complexity, imaging techniques reveal different dimensions of brain structure and function. They can be broadly classified into three groups namely [18,19]:

- Structural Neuroimaging
- Functional Neuroimaging
- Molecular Neuroimaging

Structural neuroimaging

Magnetic Resonance Imaging (sMRI), Computed Tomography (CT) and Diffusion Tensor Imaging (DTI) are some of the prominent structural neuroimaging techniques. Structural MRI is widely used to examine the shape, size and structural alterations in the brain regions [20,21]. DTI is an advanced MR technique that helps in understanding structural connectivity between brain regions [22,23]. These techniques primarily help in observable indicators such as "tissue damage" or loss of brain regions as well as measurable indicators such as white or gray matter changes and morphological changes such as cortical thinning [24,25]. These indicators are collectively classified as neuroimaging biomarkers as they are quantitative tracers of the disease progression. Some important neuroimaging biomarkers are listed below:

Atrophy

Brain atrophy is one of the most prominent neuroimaging biomarker for AD. Atrophy refers to the loss of nerves and tissue, which ultimately results in the shrinkage of the brain [26,27]. It has been previously estimated that whole brain atrophy affects 2% of AD patients while the rate of atrophy in normal ageing does not exceed beyond 0.7% per year [28]. According to Frisoni, et al., the earliest MRI based atrophic changes can be detected in entorhinal cortex, hippocampus and cingulate cortex resulting in early memory dysfunction [29,30].

Cortical thinning

Many histopathological studies have proposed that AD are often related to damage of specific cortical layers such as neocortex and entorhinal cortex [31,32]. The latest MRI techniques still are not capable of examining individual layers of the cortex. However, there are many semi-automated surface reconstruction tools such as FreeSurfer, 3D MPRAGE that aid in cortex examination [33-35].

Fractional anisotropy

DTI techniques tracts the water diffusion in various tissues which provide vital information. They help in measuring the structure of white matter as well as fiber connectivity within and across brain regions [36-38]. Fractional anisotropy is a numerical measure of fiber integrity. This index is sensitive enough to detect the white matter degradation in aging and other neurodegenerative diseases [39-41] (Figure 1).

Correspondence to: 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,

Received: January 04, 2017; **Accepted:** January 25, 2017; **Published:** January 28, 2017

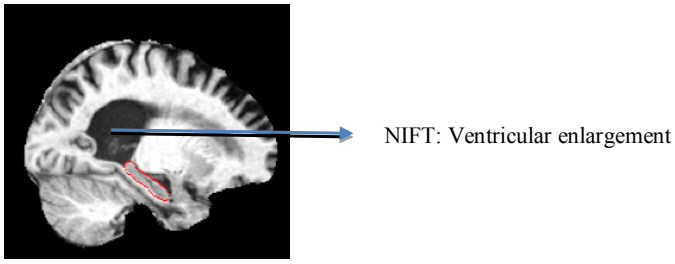


Figure 1. T13D MP-RAGE scan of structural neuroimaging**.** This figure illustrates the image scan of an AD patient with ventricular enlargement annotated using NIFT terminology.

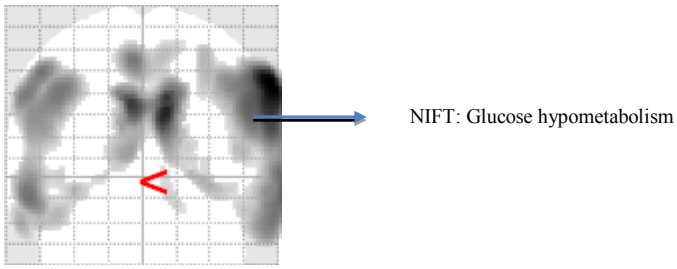


Figure 2. Functional Neuroimaging using FDG [18] PET. This figure represents an AD patient scan with excessive excessive glucose hypometabolism.

Functional neuroimaging

Functional neuroimaging techniques help in determining the severity of brain injury which reflects on the cognitive and behavioral changes amongst patients [42,43]. The most commonly used functional neuroimaging techniques include functional MRI, Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), Electroencephalography (EEG) and Transcranial Magnetic Simulation (TMS) [44,45] and MR Spectroscopy (MRS).

MRS is a widely used non-invasive imaging technique that helps in measuring the metabolites found in brain tissues. It also facilitates in measuring the chemical composition of tissues such as myo-inositol, choline, n-acetyl aspartate as well as choline. The advanced MRS imaging techniques help in identifying patients much ahead of the clinical onset of AD [46,47].

Brain glucose metabolism

Recent advancements in functional imaging studies have contributed significantly to identification of patterns amongst patients, who are at the risk of developing AD [48,49]. The earliest PET imaging abased studies were used to detect altering glucose metabolic changes amongst patients who were at a genetic risk of developing AD [50-52]. PET-based radioisotopes such as oxygen (0^{15}) aid in tracing changes in cerebral blood flow which are often caused due to increased neuronal activity [53-55]. Similarly, [18 F] fludeoxyglucose-positron emission tomography (FDG-PET) detects bilateral temporoparietal hypometabolism [56-58]. They have been widely used as a diagnostic differential biomarker discriminating between patients with AD dementia and vascular dementia [59,60]. Another radioisotope based biomarker that is widely used in diagnostic studies is C-labeled Pittsburgh Compound-B ([(11)C]PIB). The increased binding potential of PiB was found to be common amongst MCI patients whereas decreased FDG uptake was observed only with patients with AD, thus serving a crucial diagnostic biomarker [61,62] (Figure 2).

Perfusion

Imaging techniques such as SPECT and DTI enable early detection of hypoperfusion in the white matter and cortex [63,64]. Abnormal cerebral perfusion are clear indicators of diagnostic transition from MCI to AD [65,66]. Borroni and Chao et al., has demonstrated patterns of hypoperfusion in parietal, temporal and posterior cingulate cortex in all those patients who are progressing from MCI to AD [67,68]. Another study performed by Caroli et al., compared three diagnostic groups namely CN, MCI and AD. The outcome of this study reported that hippocampal hypoperfusion pattern was found across patients with amnestic MCI in transition to AD [69] (Figure 3).

Emerging combinatorial biomarkers for AD

Clinical neuroimaging biomarkers are useful resources for AD diagnosis. However, the characteristics of these imaging biomarkers are not yet adequate for diagnosis of patients at an individual level. This is largely due to the lack of longitudinal imaging data [70,71]. Combining known genetic biomarkers with imaging data could improve the prediction pattern across all patients [72-74]. Neuroimaging genetics is an emerging field in which quantitative phenotypic features from brain imaging are used as readout to inspect the role of genetic variation in brain function [75,76].

Large scale GWAS studies have contributed to the identification of many risk mutations associated with AD such as CLU, PICALM, BIN1, CR1 and so on [77-79]. These studies have created a substantial shift in the mundane AD detection through standard cognitive tests. Of all the above mentioned genes, CLU is the most significant gene used in combinatorial imaging analysis. The risk variant rs11136000 have been associated with reduction in hippocampal volume in patients with Late Onset Alzheimer Disease (LOAD) [80-82]. Apart from CLU, the risk variant rs541458 of PICALM was found to be associated with CSF Abeta 42 levels [83-85]. Similarly, large scale initiatives across the globe have already started investing in the direction of combining genetic and imaging derived biomarkers for better AD diagnosis (Table 1).

Large scale initiatives on neuroimaging and genetics

Here, we summarize the various initiatives that are focusing on integrating multi-scale data such as imaging and genetics for efficient diagnosis and treatment.

ADNI

ADNI is considered as one of the biggest ongoing multicenter study for developing longitudinal clinical, imaging, genetic and neuropsychological biomarkers for early detection of AD. The initial phase (ADNI-1) study had the greatest enrollment of participants comprising of 400 early MCI subjects, 200 AD and 200 Controls. Owing to its success, the study was further extended into the next

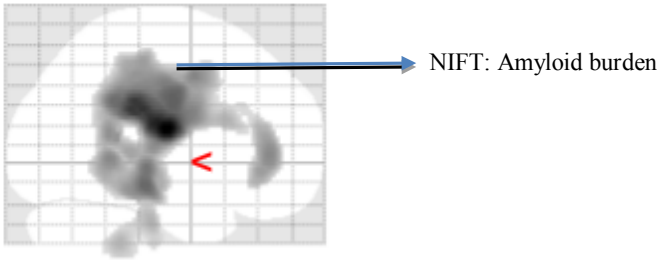


Figure 3. Molecular neuroimaging using [18F] AV 45 PET. This figure displays the amyloid burden in coronal, sagittal regions of AD patient.

Table 1. Represents a sample of cohort-based studies done using combinations of biomarkers for early AD detection.

Study type	Cohort	Snps	Imaging readouts	Outcome
Voineskos, *et al.* [86]	Philadelphia Neurodevelopmental cohort	rs12148337	White matter fractional anisotropy	The mutation had a polygenic risk score with white matter FA in schizophrenic population
Louwersheimer, *et al.* [87]	Amsterdam Dementia Cohort	rs2070045-G (SORL1)	Hippocampal atrophy	SORL1 SNP rs2070045-G allele was related to CSF-tau and hippocampal atrophy, 2 endophenotype markers of AD, suggesting that SORL1 may be implicated in the downstream pathology in AD.
Benussi, *et al.* [88]	Brescia Cohort	Leu271LeufsX10 (PGRN)	Multiple System Atrophy	PGRN mutations were shown in familial FTLD, 75% in Corticobasal syndrome
Morris, *et al.* [89]	Cohort of 355 stroke survivors	rs1799983, p.Asp298Glu	Cerebral perfusion	The presence of TT genotype increased risk of incident dementia compared with GG genotype; hazard ratio, 3.14 (95% confidence interval, 1.64-5.99; p = 0.001).
Schuur M, *et al.* [90]	Dutch family based cohort	rs1699102, rs3824968, rs2282649, rs1010159	Microbleeds	The association of SORL1 with microbleeds suggests that the amyloid cascade is involved in the aetiology of microbleeds in populations with hypertension.
Inkster, *et al.* [91]	AD cohort	rs10868366	Gray matter volume	The greater effect size in AD patients also suggests that the GG genotype could be a risk factor for the expression of cognitive deficits in AD.
Lyall, *et al.* [92]	Lothian Birth Cohort	rs10524523	Hippocampal volume	
Assareh, *et al.* [93]	Longitudinal Sydney Older Persons Study	rs4935774-T, rs2298813-G, rs1133174-G	Hippocampal atrophy	The most common haplotype (H1), comprising rs4935774-T, rs2298813-G, and rs1133174-G alleles (T/G/G) was associated with whole brain atrophy in both males and females (p=0.012 & p=0.013; respectively).
Oliveira-Filho, *et al.* [94]	Boston Cohort	rs20417	White matter hyperintensity volume	rs20417 polymorphism was associated with increased WMHv (P = .037),not cardioembolic stroke patients.

phase (ADNI-2) with additional 550 participants. This study aimed at developing a standardized protocol for data integration and collection for MRI, PET and CSF biomarkers in a global environment [95,96]. The outcome of this study produced interesting hypotheses which went beyond conventional understanding of the AD pathology. One of the earlier studies demonstrated that image derived biomarkers such as "atrophy" and "hypometabolism" exhibited a pattern based on the disease progression and severity [97,98]. Many successive studies also demonstrated the importance of CSF biomarkers, PET based biomarkers as early indicators of pre-clinical AD [99-101]. Another sister initiative of ADNI is called ADNI Genetics Core, which provides the possibility for researchers to estimate the genetic alterations using imaging features for understanding disease progression over time [102-104].

The European Alzheimer's disease Neuroimaging Initiative (E-ADNI)

The overall goal of the E-ADNI initiative was to apply the standardized protocol of collecting images, genetics, and clinical as well as psychological data by adapting the European Centers of the Alzheimer 's disease Consortium (EADC). This initiative was propelled to encourage the academic EADC centers to adopt the ADNI protocol for enrolling participants [105,106].

The Italian Alzheimer's Disease Neuroimaging Initiative (I-ADNI)

The I-ADNI initiative was launched in succession to US-ADNI study for validating the acquisition and processing protocol of structural MRI scans obtained from different clinics across Italy by following the procedure from the original ADNI study [107,108].

The Australian Imaging Biomarkers and Lifestyle Study of Aging (AIBL)

The AIBL (https://aibl.csiro.au/about/) initiative consists of 1,200 Australian participants who were longitudinally assessed for over 5 years. This study was launched in 2006 to identify biomarkers, cognitive assessments, genotype, biomarkers such as APOE, social and health factors for monitoring AD progression and early AD treatment. The AIBL initiative has given rise to lot of insights such as AD patients are prone to be more anemic than patients with MCI [109,110]. Participants enrolled in this initiative are continuously assessed every 18 months for any clinical indication of the disease.

EPAD

EPAD (http://ep-ad.org/) stands for European Prevention of Alzheimer's Dementia Consortium. It is a major European initiative for developing systematic and flexible approaches to clinical trials of drugs for preventing Alzheimer's dementia. The adaptive trial design in EPAD promises a faster and low cost drug production in the market. The imaging protocol of EPAD is adapted from the AMYPAD initiative which brings together the academic and private research groups for PET based studies to explore amyloid-beta as a therapeutic marker for AD [111,112].

AMYPAD

AMYPAD (http://www.amypad.eu/) stands for Amyloid Imaging to Prevent AD. This project was initiated to investigate the beta-amyloid biology through PET scans from pre-symptomatic population as a diagnostic and therapeutic biomarker for AD. The AMYPAD project is funded by the Innovative Medicine Initiative (IMI) program

and will run initially over 5 years. In the course of this project, patients susceptible to AD will be scanned for beta-amyloid through PET imaging. The initiative aims at improving the diagnostic standards for AD treatment and prevention (http://www.alzheimer-europe.org/News/EU-projects/Thursday-17-December-2015-AMYPAD-project-progresses-to-second-stage-of-applications-for-IMI2-Call-5).

PPMI

The Parkinson's Progressive Markers Initiative (PPMI) (http://www.ppmi-info.org/) is an observational longitudinal clinical study designed for examining patients with Parkinson's Disease (PD), healthy controls and also participants who have higher chances of developing PD. This study comprises of 1000 participants examined regularly over 8 years and the patients are enrolled in 33 clinical sites in the US, in Europe, and in Asia. The PPMI data resource comprises of clinical and behavioral assessments, imaging data and biospecimen such as CSF, DNA, RNA, plasma, urine and cell line samples. PPMI is funded by the Michael J Fox Foundation in collaboration with 18 biotech and pharmaceutical companies [113].

ENIGMA

ENIGMA (http://enigma.ini.usc.edu/) stands for Enhancing NeuroImaging Genetics Through Meta-Analysis. This consortium is an effort towards bringing researchers from diverse domains such as imaging genomics, neurology and psychiatry together to understand brain structure and function through MRI, DTI, fMRI, genetic as well as patient data. This study has so far analyzed 12,826 subjects. The preliminary project of ENIGMA was to identify common genetic variants in hippocampal or intracranial volume using Genome Wide Association Studies (GWAS). ENIGMA2 was the next project to explore genetic variants associated with subcortical volumes and ENIGMA-DTI was designed to explore genetic variants associated with white matter microstructures. Apart from meta-analysis based studies, the consortia are also focusing on understanding, how psychiatric conditions such as schizophrenia, bipolar disorder, depression affect brain functionality [114,115].

NeuroImage

NeuroImage (http://www.neuroimage.nl/) is an International Multiscale Attention-Deficit/Hyperactivity Disorder (ADHD) Genetics Initiative (IMAGE) funded by the National Institute of Mental Health. The goal of the study is to gather and analyze endophenotypic, phenotypic and genetic information about ADHD. This study is based on a collection of 5,578 subjects from 8 European countries. In the course of this project, structural and functional MRI scans are performed on patients, along with neuropsychological assessments and GWAS analysis in order to detect functional abnormalities underlying ADHD [116,117].

Initiatives such as ADNI and PPMI have largely invested in systematically harvesting genetic and imaging data. Studies like ADNI and PPMI form the basis for the association of imaging readouts with genetic variation information and may facilitate the generation of hypotheses about mechanistic links between genes and imaging features.

Mining links between neuroimaging readouts and molecular processes from literature

High-throughput imaging technologies have been employed to understand the molecular mechanisms underlying clinical conditions. Such efforts have led to the identification of novel biomarkers for all disease domains, especially AD [118,119]. However, the rapid growth of the literature around these combinatorial studies has made it increasingly difficult to aggregate and mine the reported findings [120]. Obviously, new technologies enabling automated text processing ("text-mining") may help to retrieve relevant documents and to extract relevant knowledge from text.

Ontologies and terminologies

One of the most efficient ways to address the challenge of unstructured information mining is with the efficient usage of ontologies and controlled terminologies. Ontologies are formal representations of knowledge that can represent entire research domains. They are helpful when concepts need to be shared across research communities in an unambiguous fashion. This is very crucial as it enables different research groups to communicate with each other without misinterpretation of the biological context [121-123]. Ontologies do also facilitate the exchange of data and knowledge between machines; they are in fact readable by both, human experts and machines. When transformed into terminologies (dictionaries), they can readily be integrated into text-mining systems and are very useful for information extraction and knowledge representation. Furthermore, ontologies bear the potential to enable automated reasoning over knowledge representations [124,125].

Existing ontologies in the field of neuroimaging

Similar to other biological domains, the field of neuroimaging research has advanced semantically by generating various terminologies and ontologies in the past. Some of the more widely recognized neuroimaging ontologies are listed below:

Quantitative Imaging Biomarker Ontology (QIBO)

QIBO ontology was developed to standardize quantitative imaging biomarkers for better therapeutic intervention. This ontology consists of 488 terms and they consist of classes such as imaging agent, imaging instrument or biological intervention. QIBO represents concepts across several fields, including imaging physics and biology [126].

Magnetic Resonance Imaging Ontology (MRIO)

This ontology captures all concepts needed to describe the outcome of MRI scans. It has been designed to overcome the heterogeneity in MRI readouts. The authors mainly capture measured data coming from T1, T2, tissue as well as other factors, such as temperature. The MRIO ontology focusses mainly on two MRI representations namely MRI simulators and DICOM images and conceptualize all possible terms that can be observed using these scanned images [127,128].

NeuroLog

The NeuroLog consortium was established in the year 2006 for sharing and reusing data and tools for neuroimaging studies. This software architecture aids in efficient integration of neuroimaging data and tools from various neuroimaging research centers. This consortium also takes charge of the autonomous data management from each center to maintain the confidentiality of the neuroimaging data. Furthermore, the usage of semantically annotated tools inbuilt in the system architecture provides better standardization of neuroimaging datasets and therefore offers better accessibility through the federated schema based ontology [129,130].

NeuroImage Feature Terminology (NIFT)

Although there are so many ontologies established in the area of neuro-imaging, there is still a lack of a terminology which facilitates a systematic representation and retrieval of measured indices with high relevance for neurodegenerative diseases. All the existing ontologies represent what the imaging scan capture, but they do not contain concepts that link imaging readouts to disease pathology. Motivated by the apparent need for such a terminology, we have developed NIFT, the "neuro-image feature terminology". NIFT represents a wide spectrum of terms linked to radiological, neuropsychological as well as measured indices highly relevant to neurodegenerative diseases (e.g. AD and PD) [131]. The NIFT terminology comprises highly generic concepts describing common neuroimaging features, but at the same time it is very specific and represents disease-centric pathological measures used in imaging scans in the domain of Alzheimer 's disease and Parkinsonism. NIFT can act as a potential resource to capture molecular as well as clinical readouts, which are crucial in bridging these two domains as well as retrieving relevant documents which can be further used in a multi-layered disease models. As such, NIFT is well suited to support the identification of novel mechanisms underlying the etiology of AD and PD.

Retrieval of relevant publications using the nift terminology

The main purpose for developing ontologies and terminologies is to retrieve relevant publications and automatically extract relevant information from the literature. To enable specific retrieval and information extraction in the imaging domain, we integrated the NIFT terminology into our in-house text-mining system SCAIView [132,133]. SCAIView was developed at Fraunhofer SCAI to enable biologists and clinical researchers to perform semantic search and information extraction from the scientific literature. A free version of this literature mining environment, SCAIView academia, allows free access to the semantically annotated PubMed abstracts. For PubMed Central (PMC) full text publications, SCAIView allows a full-text search as well. We have integrated NIFT in SCAIView and used the system to systematically retrieve relevant documents containing useful information on imaging readouts linked to molecular entities. The resulting literature corpus was then used for mechanistic modeling purposes.

Mechanistic modeling of neuroimaging indices

We wanted to understand the significance of a measured index obtained from imaging techniques and their association with clinical tests to improve the prediction an underlying neurodegenerative disease, in this case, AD. For this, we performed an optimized search query using our literature-mining environment SCAIView.

We used the query "[Neuroimaging Feature]) AND [MeSH Disease: "Alzheimer Disease"]) AND [Alzheimer Ontology Node: "Evaluation"]) AND [BRCO]) AND [PTS]) AND [Organism: "Homo sapiens"]" to retrieve relevant publications that comprises disease-specific terms, brain region and cell-type information (BRCO) and that comprise pathway mentions (PTS). The Alzheimer Ontology (ADO) concept "evaluation" provides a wide spectrum of entities that describe various clinical tests that are significant for diagnosing AD. Once the articles were retrieved, we tried to model them in order to identify underlying the molecular mechanisms.

Mechanistic modeling of neuroimaging features with molecular pointers

One major motivation to develop the NIFT terminology was to support the generation of cause-and-effect models in the area of neurodegenerative diseases. With the integration of imaging features in cause-and-effect models, we hope to bridge between the molecular level (genome, pathways) and the macroscopic anatomical level of brain structures such as brain regions and the entire organ.

Using the query described above, we generated a literature corpus highly enriched for mentions of interesting imaging features together with interesting molecular processes. One of the resulting models that link imaging features to the molecular pathophysiology of AD deals with the influence of cerebral blood flow on cognitive impairment in AD. The overall workflow applied is shown in Figure 4.

NIFT application example

Hypothetical model for linking high-level cerebral blood flow with molecular processes:

The scientific community has long been interested in the vascular biology, in which the human physiology is represented as large and small blood vessels which might play a role in AD progression [134,135]. Although clinical studies conducted on AD patients reveal substantial evidences of vascular lesion being the biggest factor of AD, the fundamental understanding of the molecular mechanism behind that remains unexplained [136,137]. Therefore, here we establish our first hypothetical model that links high level complex biology such as cerebral blood flow with molecular processes. This model is highly putative due to the lack of experimental validation and lack of clinical resources to support the hypothesis.

AD is highly diverse and complex in terms of the various cellular and molecular players that together result in the disease pathology. Apart from the molecular deposits such as plaques and tangles, increasing supporting evidences on the role of vascular abnormalities in AD pathology, so much so that these co-morbid conditions are classified under the term "vascular dementia" [138-140]. The links between vascular lesions and cognition impairment are based on observations that have been captured using advanced neuroimaging techniques such as SPECT [141,142]. By using radioisotopic tracers, depletion of blood flow can be traced by reduced glucose consumption in a particular brain region [143,144]. Apart from SPECT, MRI tensors

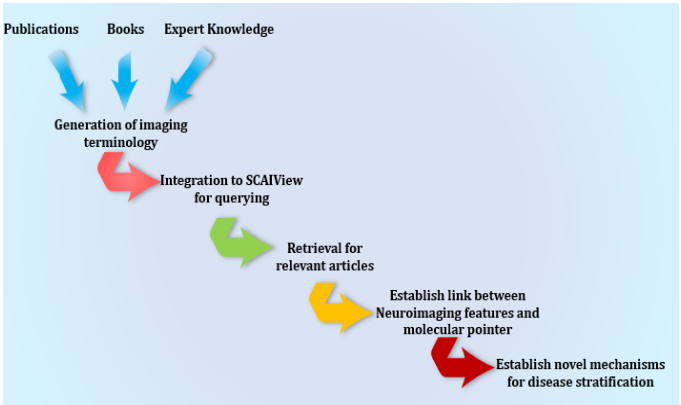


Figure 4. Schematic representation of the workflow to extract links between imaging features and molecular mechanisms in a disease context.

are constantly tuned to detect early neoplasms and altered blood flow detection with high-resolution quality [145,146].

Hypothetical mechanism for cerebral blood flow in AD

SIRT1 stands for "Silient Information Regulator 2 homolog 1". In general, their role is to maintain cellular functions and promote longevity of the cells in humans as well as other model organisms [147,148]. Sirtuins have been reported to protect the brain from infarction by regulating the blood flow to all parts of the brain, especially the cerebral region [149-151]. In normal conditions, SIRT1 has been reported to play a protective role by enhancing the non-amyloidogenic cleavage of amyloid-beta protein (APP) through NF-kb inhibition. The inhibition of NF-kb contributes to the clearance of amyloid plaques from the brain [152,153]. However, in case of AD, SIRT1 genes are reported to be under expressed which in turn activates the accumulation of amyloid beta in cerebral cortex through NF-Kb activation. The accumulation of APP in the cerebral region could further lead to the depletion of nutrients such as oxygen from the blood, resulting in the inhibition of cerebral blood flow. Lack of oxygen and other nutrients to the brain, various mental and psychiatric abnormalities and could lead to cognitive impairment [154,155].

Also, we hypothesize that the overexpression of SIRT1 co-activates a regulator, which transcribes ADAM10 [156-158]. This could trigger ADAM10 to partially compete with the gamma-secretase for APP fragment resulting in the activation of Notch signaling pathway which is well-known for neuronal repair [159-161]. However, in case of AD ADAM mutant Q170H and R181G does not compete with alpha-secretase, therefore the beta-secretases accumulate in the brain resulting in impaired cerebral blood flow [162-164].

Another plausible mechanism of reduced cerebral blood flow is due to APOE activity. Increased expression of APOE also facilitates the molecular interaction between amyloid beta and Butyrylcholineesterase (BCHE) gene which results in the formation of a complex BCHE-Abeta-APOE (BaβA) complex [165-167]. This complex alters the structure of BCHE which accelerates the catalytic activity of the enzyme. This results in the formation of amyloid plaques [168-170] as seen Figure 5. Increased expression of APOE also disrupts the neuronal activity in the hippocampus resulting in atrophy. Hippocampal atrophy is also one of the causative factor of cognitive decline in AD [171-173].

Apart from the well-known genes of AD, recently, PICALM gene has been emerging as a potential AD candidate. PICALM plays a crucial role in intracellular trafficking of endothelial proteins resulting in endocytosis. The protective allele of PICALM, rs3851179 facilitates the amyloid beta clearance through endocytosis [174-176]. LRP1 is another crucial protein whose major function is cholesterol transport and transcytosis of various molecules including amyloid beta across the BBB [177-179]. As PICALM plays a major role in the internalization of the endothelial proteins, it also internalizes the sLRP1 and amyloid-beta complex by trafficking through two other proteins Rab5 and Rab11. These further results in amyloid transcytosis and clearance from entering the BBB [180-182]. Also, LRP1 activates another protein called GLUT1 which is another major glucose transporter across the BBB [183-185]. During normal conditions, there is a free flow of glucose and other nutrients across BBB. However, during AD, GLUT function is altered by Gly286Asp resulting in inhibition of glucose metabolism [186-188].

Here, we have demonstrated a hypothetical mechanism around cerebral blood flow in AD. We call this model as "putative" and

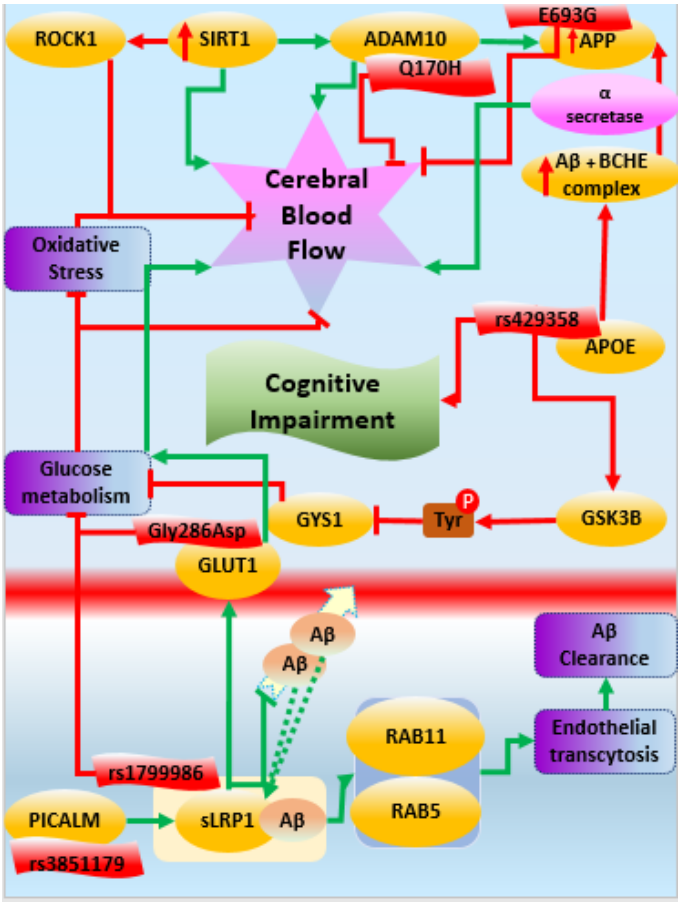


Figure 5. Mechanistic modeling of cerebral blood flow in AD.

"hypothetical" because they lack individual causal proof and substantial experimental validation. The overall workflow of the altered regulation of cerebral blood flow can be seen in Figure 5.

References

1. Iqbal K, Grundke-Iqbal I (2010) Alzheimer's disease, a multifactorial disorder seeking multitherapies. *Alzheimers Dement* 6: 420-424. [crossref]
2. Carmo Carreiras M, Mendes E, Jesus Perry M, Paula Francisco A, Marco-Contelles J (2013) The multifactorial nature of Alzheimer's disease for developing potential therapeutics. *Current topics in medicinal chemistry* 13: 1745-1770. [crossref]
3. Zekanowski C, Styczynska M, Peplonska,B, Gabryelewicz T, Religa D, et al. (2003) Mutations in presenilin 1, presenilin 2 and amyloid precursor protein genes in patients with early-onset Alzheimer's disease in Poland. *Experimental neurology* 184: 991-996. [crossref]
4. Ryan NS, Rossor MN (2010) Correlating familial Alzheimer's disease gene mutations with clinical phenotype. *Biomark Med* 4: 99-112. [crossref]
5. Weggen S, Beher D (2012) Molecular consequences of amyloid precursor protein and presenilin mutations causing autosomal-dominant Alzheimer's disease. *Alzheimer's research & therapy* 4: 1. [crossref]
6. Combarros O, Alvarez-Arcaya A, Sanchez-Guerra M, Infante J, Berciano J (2002) Candidate gene association studies in sporadic Alzheimer's disease. *Dementia and geriatric cognitive disorders* 14: 41-54. [crossref]
7. Bekris LM, Yu CE, Bird TD, Tsuang DW (2010). Review article: genetics of Alzheimer disease. *Journal of geriatric psychiatry and neurology* 23: 213-227. [crossref]
8. Makovac E, Cercignani M, Serra L, Torso M, Spanò B, et al. (2016) Brain Connectivity Changes in Autosomal Recessive Parkinson Disease: A Model for the Sporadic Form. *PloS one* 11: e0163980. [crossref]
9. Antila K, Lötjönen,J, Thurfjell L, Laine J, Massimini M, et al. (2013) The PredictAD

project: development of novel biomarkers and analysis software for early diagnosis of the Alzheimer's disease. *Interface focus* 3: 20120072. [crossref]

10. Galvin JE (2012) Optimizing diagnosis and management in mild-to-moderate Alzheimer's disease. *Neurodegener Dis Manag* 2: 291-304. [crossref]

11. Thaipisuttikul P, Galvin JE (2012) Use of medical foods and nutritional approaches in the treatment of Alzheimer's disease. *Clin Pract (Lond)* 9: 199-209. [crossref]

12. Barrett JJ, Haley WE, Harrell LE, Powers RE (1997) Knowledge about Alzheimer disease among primary care physicians, psychologists, nurses, and social workers. *Alzheimer Disease & Associated Disorders* 11: 99-106. [crossref]

13. Hunter CA, Kirson NY, Desai U, Cummings AKG, Faries DE, et al. (2015) Medical costs of Alzheimer's disease misdiagnosis among US Medicare beneficiaries. *Alzheimer's & Dementia* 11: 887-895. [crossref]

14. Yiannopoulou KG, Papageorgiou SG (2013) Current and future treatments for Alzheimer's disease. *Ther Adv Neurol Disord* 6: 19-33. [crossref]

15. Galimberti D, Scarpini E (2011) Disease-modifying treatments for Alzheimer's disease. *Ther Adv Neurol Disord* 4: 203-216. [crossref]

16. Nestor PJ, Scheltens P, Hodges JR (2004) Advances in the early detection of Alzheimer's disease. *Nat Med* 10 Suppl: S34-41. [crossref]

17. Small GW (2000) Early diagnosis of Alzheimer's disease: update on combining genetic and brain-imaging measures. *Dialogues Clin Neurosci* 2: 241-246. [crossref]

18. Rudin M (2009) Noninvasive structural, functional, and molecular imaging in drug development. *Current opinion in chemical biology* 13: 360-371. [crossref]

19. Saykin A, Shen L, Risacher S, Swaminathan S, Kim S, et al. (2012) Structural, functional and molecular imaging as intermediate phenotypes for studies of candidate genes, pathways and GWAS in Alzheimer's disease. Alzheimer's & Dementia: *The Journal of the Alzheimer's Association* 8: 232. [crossref]

20. Fallon N, Alghamdi J, Chiu Y, Sluming V, Nurmikko T, et al. (2013) Structural alterations in brainstem of fibromyalgia syndrome patients correlate with sensitivity to mechanical pressure. *NeuroImage* 3: 163-170. [crossref]

21. Alexander-Bloch A, Giedd JN, Bullmore E (2013) Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience* 14: 322-336. [crossref]

22. Mori S, Zhang J (2006) Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron* 51: 527-539. [crossref]

23. Kalinosky BT, Schindler-Ivens S, Schmit BD (2013) White matter structural connectivity is associated with sensorimotor function in stroke survivors. *NeuroImage* 2: 767-781. [crossref]

24. Kharitonova M, Martin RE, Gabrieli JD, Sheridan MA (2013) Cortical gray-matter thinning is associated with age-related improvements on executive function tasks. *Developmental cognitive neuroscience* 6: 61-71. [crossref]

25. Ouyang X, Chen K, Yao L, Hu B, Wu X, et al. (2015) Simultaneous changes in gray matter volume and white matter fractional anisotropy in Alzheimer's disease revealed by multimodal CCA and joint ICA. *Neuroscience* 301: 553-562. [crossref]

26. Peters R (2006) Ageing and the brain. *Postgrad Med J* 82: 84-88. [crossref]

27. Palop JJ, Mucke L (2016) Network abnormalities and interneuron dysfunction in Alzheimer disease. *Nature Reviews Neuroscience* 17: 777–792. [crossref]

28. Sluimer JD, van der Flier WM, Karas GB, Fox NC, Scheltens P, et al. (2008) Whole-brain atrophy rate and cognitive decline: longitudinal MR study of memory clinic patients 1. *Radiology* 248: 590-598. [crossref]

29. Frisoni GB, Ganzola R, Canu E, Rüb U, Pizzini FB, et al. (2008) Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. *Brain* 131: 3266-3276. [crossref]

30. Frisoni GB, Laakso MP, Beltramello A, Geroldi C, Bianchetti A, et al. (1999) Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease. *Neurology* 52: 91-100. [crossref]

31. Kersaitis C, Halliday GM, Kril JJ (2004) Regional and cellular pathology in frontotemporal dementia: relationship to stage of disease in cases with and without Pick bodies. *Acta Neuropathol* 108: 515-523. [crossref]

32. Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT (2011) Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med* 1: a006189. [crossref]

33. Du AT, Schuff N, Kramer JH, Rosen HJ, Gorno-Tempini ML, et al. (2007) Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain* 130: 1159-1166. [crossref]

34. Lüsebrink F, Wollrab A, Speck O (2013) Cortical thickness determination of the human brain using high resolution 3T and 7T MRI data. *Neuroimage* 70: 122-131. [crossref]

35. Redolfi A, Manset D, Barkhof F, Wahlund LO, Glatard T, et al. (2015) Head-to-head comparison of two popular cortical thickness extraction algorithms: a cross-sectional and longitudinal study. *PloS one* 10: e0117692. [crossref]

36. Fonteijn HM, Norris DG, Verstraten FA (2008) Exploring the anatomical basis of effective connectivity models with DTI-based fiber tractography. *Journal of Biomedical Imaging* 2008: 3. [crossref]

37. Soldner J, Meindl T, Koch W, Bokde AL, Reiser MF, et al. (2012) [Structural and functional neuronal connectivity in Alzheimer's disease: a combined DTI and fMRI study]. *Nervenarzt* 83: 878-887. [crossref]

38. Acosta-Cabronero J, Nestor PJ (2014) Diffusion tensor imaging in Alzheimer's disease: insights into the limbic-diencephalic network and methodological considerations. *Frontiers in aging neuroscience* 6: 266. [crossref]

39. Zhang Y, Schuff N, Jahng GH, Bayne W, Mori S, et al. (2007) Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease. *Neurology* 68: 13-19. [crossref]

40. Bennett IJ, Madden DJ, Vaidya CJ, Howard DV, Howard JH (2010) Age-related differences in multiple measures of white matter integrity: A diffusion tensor imaging study of healthy aging. *Human brain mapping* 31: 378-390. [crossref]

41. Welton T, Kent D, Constantinescu CS, Auer DP, Dineen RA (2014) Functionally relevant white matter degradation in multiple sclerosis: a tract-based spatial meta-analysis. *Radiology* 275: 89-96. [crossref]

42. Ng TS, Lin AP, Koerte IK, Pasternak O, Liao H, et al. (2014) Neuroimaging in repetitive brain trauma. *Alzheimers Res Ther* 6: 10. [crossref]

43. Beauregard M (2014) Functional neuroimaging studies of the effects of psychotherapy. *Dialogues Clin Neurosci* 16: 75-81. [crossref]

44. Johnson KA, Fox NC, Sperling RA, Klunk WE (2012) Brain imaging in Alzheimer disease. *Cold Spring Harb Perspect Med* 2: a006213. [crossref]

45. Niethammer M, Feigin A, Eidelberg D (2012) Functional neuroimaging in Parkinson's disease. *Cold Spring Harb Perspect Med* 2: a009274. [crossref]

46. Chen JG, Charles HC, Barboriak D, Doraiswamy PM (2000) Magnetic resonance spectroscopy in Alzheimer's disease: focus on N□acetylaspartate. *Acta Neurologica Scandinavica 102*: 20-26. [crossref]

47. Mandal PK (2007) Magnetic resonance spectroscopy (MRS) and its application in Alzheimer's disease. *Concepts in Magnetic Resonance Part A* 30: 40-64 [crossref]

48. Ferreira LK, Busatto GF (2011) Neuroimaging in Alzheimer's disease: current role in clinical practice and potential future applications. *Clinics (Sao Paulo)* 66 Suppl 1: 19-24. [crossref]

49. Mak E, Su L, Williams GB, O'Brien JT (2014) Neuroimaging characteristics of dementia with Lewy bodies. *Alzheimers Res Ther* 6: 18. [crossref]

50. Small GW (1996) Neuroimaging and genetic assessment for early diagnosis of Alzheimer's disease. *J Clin Psychiatry* 57 Suppl 14: 9-13. [crossref]

51. von Leden RE, Selwyn RG, Jaiswal S, Wilson CM, Khayrullina G, et al. (2016) 18 F-FDG-PET imaging of rat spinal cord demonstrates altered glucose uptake acutely after contusion injury. *Neuroscience letters* 621: 126-132. [crossref]

52. Schroeder FA, Chonde DB, Riley MM, Moseley CK, Granda ML, et al. (2013) FDG-PET imaging reveals local brain glucose utilization is altered by class I histone deacetylase inhibitors. *Neuroscience letters* 550: 119-124. [crossref]

53. Owler BK, Momjian S, Czosnyka Z, Czosnyka M, Péna A, et al. (2004) Normal Pressure Hydrocephalus and Cerebral Blood Flow: A PET Study of Baseline Values. *Journal of Cerebral Blood Flow & Metabolism* 24: 17-23. [crossref]

54. Grüner JM, Paamand R, Højgaard L, Law I (2011) Brain perfusion CT compared with 15 OH 2 O-PET in healthy subjects. *EJNMMI research* 1: 1. [crossref]

55. Su Y, Vlassenko A, Blazey T, Ances B, Snyder A, et al. (2014) Comparison of cerebral blood flow measurement obtained from simultaneously acquired ASL and O15-water PET. *J Nucl Med* 55: 207-207. [crossref]

56. Womack KB, Diaz-Arrastia R, Aizenstein HJ, Arnold SE, Barbas NR, et al. (2011) Temporoparietal hypometabolism in frontotemporal lobar degeneration and associated imaging diagnostic errors. *Archives of Neurology* 68: 329-337. [crossref]

57. Newberg AB, Arnold SE, Wintering N, Rovner BW, Alavi A (2012) Initial clinical

comparison of 18F-florbetapir and 18F-FDG PET in patients with Alzheimer disease and controls. *J Nucl Med* 53: 902-907. [crossref]

58. Marcus C, Mena E, Subramaniam RM (2014) Brain PET in the diagnosis of Alzheimer's disease. *Clin Nucl Med* 39: e413-422. [crossref]

59. Albin RL, Koeppe RA, Burke JF, Giordani B, Kilbourn MR, et al. (2010) Comparing Fludeoxyglucose F18–PET Assessment of Regional Cerebral Glucose Metabolism and [11C] Dihydrotetrabenazine–PET in Evaluation of Early Dementia and Mild Cognitive Impairment. *Archives of neurology* 67: 440-446. [crossref]

60. Ruan Q, D'Onofrio G, Sancarlo D, Bao Z, Greco A, et al. (2016) Potential neuroimaging biomarkers of pathologic brain changes in Mild Cognitive Impairment and Alzheimer's disease: a systematic review. *BMC geriatrics* 16: 1. [crossref]

61. Mosconi L, Berti V, Glodzik L, Pupi A, De Santi S, et al. (2010) Pre-clinical detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging. *J Alzheimers Dis* 20: 843-854. [crossref]

62. Cohen AD, Klunk WE (2014) Early detection of Alzheimer's disease using PiB and FDG PET. *Neurobiol Dis* 72 Pt A: 117-122. [crossref]

63. Risacher SL, Saykin A J (2013) Neuroimaging biomarkers of neurodegenerative diseases and dementia. *Semin Neurol* 33: 386-416. [crossref]

64. Ruan Q, D'Onofrio G, Sancarlo D, Bao Z, Greco A, et al. (2016) Potential neuroimaging biomarkers of pathologic brain changes in Mild Cognitive Impairment and Alzheimer's disease: a systematic review. *BMC geriatrics* 16: 1. [crossref]

65. Luckhaus C, Cohnen M, Flüb MO, Jänner M, Grass-Kapanke B, et al. (2010) The relation of regional cerebral perfusion and atrophy in mild cognitive impairment (MCI) and early Alzheimer's dementia. Psychiatry Research: *Neuroimaging* 183: 44-51. [crossref]

66. Hauser T, Schönknecht P, Thomann PA, Gerigk L, Schröder J, et al. (2013) Regional cerebral perfusion alterations in patients with mild cognitive impairment and Alzheimer disease using dynamic susceptibility contrast MRI. *Academic radiology* 20: 705-711. [crossref]

67. Borroni B, Anchisi D, Paghera B, Vicini B, Kerrouche N, et al. (2006) Combined 99mTc-ECD SPECT and neuropsychological studies in MCI for the assessment of conversion to AD. *Neurobiol Aging* 27: 24-31. [crossref]

68. Chao LL, Buckley ST, Kornak J, Schuff N, Madison C, et al. (2010) ASL perfusion MRI predicts cognitive decline and conversion from MCI to dementia. *Alzheimer disease and associated disorders* 24: 19. [crossref]

69. Caroli A, Testa C, Geroldi C, Nobili F, Barnden LR, et al. (2007) Cerebral perfusion correlates of conversion to Alzheimer's disease in amnestic mild cognitive impairment. *Journal of neurology* 254: 1698-1707. [crossref]

70. Shokouhi S, Campbell D, Brill AB, Gwirtsman HE (2016) Longitudinal Positron Emission Tomography in Preventive Alzheimer's Disease Drug Trials, Critical Barriers from Imaging Science Perspective. *Brain Pathology* 26: 664-671. [crossref]

71. Li Y, Gilmore JH, Shen D, Styner M, Lin W, et al. (2013) Multiscale adaptive generalized estimating equations for longitudinal neuroimaging data. *NeuroImage* 72: 91-105. [crossref]

72. Batmanghelich NK, Dalca AV, Sabuncu MR, Golland P (2013) Joint modeling of imaging and genetics. In International Conference on Information Processing in Medical Imaging. Springer Berlin Heidelberg 766-777. [crossref]

73. Kong D, Giovanello KS, Wang Y, Lin W, et al. (2015) Predicting Alzheimer's Disease Using Combined Imaging-Whole Genome SNP Data. *J Alzheimers Dis* 46: 695-702. [crossref]

74. Harrison TM, Bookheimer SY (2016) Neuroimaging genetic risk for Alzheimer's disease in preclinical individuals: from candidate genes to polygenic approaches. Biol Psychiatry Cogn Neurosci Neuroimaging 1: 14-23. [crossref]

75. Wang H, Nie F, Huang H, Yan J, Kim S, et al. (2012) From phenotype to genotype: an association study of longitudinal phenotypic markers to Alzheimer's disease relevant SNPs. *Bioinformatics* 28: i619-i625. [crossref]

76. Shen L, Thompson PM, Potkin SG, Bertram L, Farrer LA, et al. (2014) Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain imaging and behavior* 8: 183-207. [crossref]

77. Tan L1, Yu JT, Zhang W, Wu ZC, Zhang Q, et al. (2013) Association of GWAS-linked loci with late-onset Alzheimer's disease in a northern Han Chinese population. *Alzheimers Dement* 9: 546-553. [crossref]

78. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, et al. (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature genetics* 41: 1088-1093. [crossref]

79. Shen L, Jia J (2016) An Overview of Genome-Wide Association Studies in Alzheimer's Disease. *Neurosci Bull* 32: 183-190. [crossref]

80. Yang X, Li J (2016) Impact of PICALM and CLU on hippocampal degeneration. *Hum Brain Mapp* 37: 2419-2430. [crossref]

81. Erk S, Meyer-Lindenberg A, Opitz von Boberfeld C, Esslinger C, Schnell K, et al. (2011) Hippocampal function in healthy carriers of the CLU Alzheimer's disease risk variant. *J Neurosci* 31: 18180-18184. [crossref]

82. DiBattista AM, Stevens BW, Rebeck GW, Green AE (2014) Two Alzheimer's disease risk genes increase entorhinal cortex volume in young adults. *Frontiers in human neuroscience* 8. [crossref]

83. Schjeide BMM, Schnack C, Lambert JC, Lill CM, Kirchheiner J, et al. (2011) The role of clusterin, complement receptor 1, and phosphatidylinositol binding clathrin assembly protein in Alzheimer disease risk and cerebrospinal fluid biomarker levels. *Archives of general psychiatry* 68: 207-213. [crossref]

84. Lambert JC1, Zelenika D, Hiltunen M, Chouraki V, Combarros O, et al. (2011) Evidence of the association of BIN1 and PICALM with the AD risk in contrasting European populations. *Neurobiol Aging* 32: 756. [crossref]

85. Schott JM; ADNI Investigators (2012) Using CSF biomarkers to replicate genetic associations in Alzheimer's disease. *Neurobiol Aging* 33: 1486. [crossref]

86. Voineskos AN, Felsky D, Wheeler AL, Rotenberg DJ, Levesque M, et al. (2015) Limited Evidence for Association of Genome-Wide Schizophrenia Risk Variants on Cortical Neuroimaging Phenotypes. *Schizophrenia bulletin* sbv180. [crossref]

87. Louwersheimer E, Ramirez A, Cruchaga C, Becker T, Kornhuber J, et al. (2015) Influence of genetic variants in SORL1 gene on the manifestation of Alzheimer's disease. *Neurobiol Aging* 36: 1605. [crossref]

88. Benussi L, Ghidoni R, Pegoiani E, Moretti DV, Zanetti O, et al. (2009) Progranulin Leu271LeufsX10 is one of the most common FTLD and CBS associated mutations worldwide. *Neurobiology of disease* 33: 379-385. [crossref]

89. Morris CM, Ballard CG, Allan L, Rowan E, Stephens S, et al. (2011) NOS3 gene rs1799983 polymorphism and incident dementia in elderly stroke survivors. *Neurobiol Aging* 32: 554. [crossref]

90. Schuur M, Van Swieten JC, Schol-Gelok S, Ikram MA, Vernooij MW, et al. (2010) Genetic risk factors for cerebral small-vessel disease in hypertensive patients from a genetically isolated population. J Neurol Neurosurg Psychiatry 2009. [crossref]

91. Inkster B, Rao AW, Ridler K, Filippini N, Whitcher B, et al. (2012) Genetic variation in GOLM1 and prefrontal cortical volume in Alzheimer's disease. *Neurobiology of aging* 33: 457-465. [crossref]

92. Lyall DM, Royle NA, Harris SE, Bastin ME, Maniega SM, et al. (2013) Alzheimer's disease susceptibility genes APOE and TOMM40, and hippocampal volumes in the Lothian birth cohort 1936. *PloS one* 8: e80513. [crossref]

93. Assareh AA, Piguet O, Lye TC, Mather KA, Broe GA, et al. (2014) Association of SORL1 gene variants with hippocampal and cerebral atrophy and Alzheimer's disease. *Curr Alzheimer Res* 11: 558-563. [crossref]

94. Oliveira-Filho J, Ornellas AC, Zhang CR, Oliveira LM, Araújo-Santos T, et al. (2015) COX-2 rs20417 Polymorphism Is Associated with Stroke and White Matter Disease. *Journal of Stroke and Cerebrovascular Diseases* 24: 1817-1822. [crossref]

95. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, et al. (2013) The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & Dementia* 9: e111-e194. [crossref]

96. Kang JH, Korecka M, Figurski MJ, Toledo JB, Blennow K, et al. (2015) The Alzheimer's Disease Neuroimaging Initiative 2 Biomarker Core: A review of progress and plans. *Alzheimers Dement* 11: 772-791. [crossref]

97. Johnson KA, Fox NC, Sperling RA, Klunk WE (2012) Brain imaging in Alzheimer disease. *Cold Spring Harb Perspect Med* 2: a006213. [crossref]

98. Dukart J, Mueller K, Villringer A, Kherif F, Draganski B, et al. (2013) Relationship between imaging biomarkers, age, progression and symptom severity in Alzheimer's disease. *NeuroImage* 3: 84-94. [crossref]

99. Palmqvist S, Zetterberg H, Mattsson N, Johansson P; Alzheimer's Disease Neuroimaging Initiative, Minthon L, et al. (2015) Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. *Neurology* 85: 1240-1249. [crossref]

100. Mosconi L, Berti V, Glodzik L, Pupi A, De Santi S, et al. (2010) Pre-clinical

detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging. *J Alzheimers Dis* 20: 843-854. [crossref]

101. Berti V, Polito C, Lombardi G, Ferrari C, Sorbi S, et al. (2016) Rethinking on the concept of biomarkers in preclinical Alzheimer's disease. *Neurol Sci* 37: 663-672. [crossref]

102. Glahn DC, Paus T, Thompson PM (2007) Imaging genomics: mapping the influence of genetics on brain structure and function. *Human brain mapping* 28: 461-463. [crossref]

103. Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, et al. (2010) Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimer's & Dementia* 6: 265-273. [crossref]

104. Shen L, Thompson PM, Potkin SG, Bertram L, Farrer LA, et al. (2014) Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain imaging and behavior 8*: 183-207. [crossref]

105. Frisoni GB, Henneman WJ, Weiner MW, Scheltens P, Vellas B, et al. (2008) The pilot European Alzheimer's disease neuroimaging initiative of the European Alzheimer's disease consortium. *Alzheimer's & Dementia* 4: 255-264. [crossref]

106. Villemagne VL, Kim SY, Rowe CC, Iwatsubo T (2014) Imago Mundi, Imago AD, Imago ADNI. *Alzheimers Res Ther* 6: 62. [crossref]

107. Cavedo E, Babiloni C, Redolfi A, Lizio R, Chiapparini L, et al. (2012) The Italian Alzheimer's Disease Neuroimaging Initiative (IADNI). *Alzheimer's & Dementia* 8: P159. [crossref]

108. Cavedo E, Redolfi A, Angeloni F, Babiloni C, Lizio R, et al. (2014) The Italian Alzheimer's Disease Neuroimaging Initiative (I-ADNI): validation of structural MR imaging. *J Alzheimers Dis* 40: 941-952. [crossref]

109. Faux NG, Rembach A, Wiley J, Ellis KA, Ames D, et al. (2014) An anemia of Alzheimer's disease. *Mol Psychiatry* 19: 1227-1234. [crossref]

110. Burnham SC, Faux NG, Wilson W, Laws SM, Ames D, et al. (2014) A blood-based predictor for neocortical A? burden in Alzheimer's disease: results from the AIBL study. *Molecular psychiatry* 19: 519-526. [crossref]

111. Ritchie CW, Molinuevo JL, Truyen L, Satlin A, Van der Geyten S, et al. (2016) Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. *Lancet Psychiatry* 3: 179-186. [crossref]

112. Dobkin BH (2016) CURRENT OPINION Editorial introductions.

113. Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, et al. (2011) The parkinson progression marker initiative (PPMI). *Progress in neurobiology* 95: 629-635. [crossref]

114. Thompson P (2014) The Enhancing Neuroimaging Genetics Through Meta-Analysis (ENIGMA) Consortium: Investigating Brain Diseases With Imaging and Genetics in 29,000 People. 61st Annual Meeting AACAP.

115. Hibar DP, ENIGMA-Consortium (2013) ENIGMA2: genome-wide scans of subcortical brain volumes in 16,125 subjects from 28 cohorts worldwide. Organization of Human Brain Mapping, Seattle, WA.

116. Cortese S, Castellanos FX (2012) Neuroimaging of attention-deficit/hyperactivity disorder: current neuroscience-informed perspectives for clinicians. *Current psychiatry reports* 14: 568-578. [crossref]

117. von Rhein D, Mennes M, van Ewijk H, Groenman AP, Zwiers MP, et al. (2015) The NeuroIMAGE study: a prospective phenotypic, cognitive, genetic and MRI study in children with attention-deficit/hyperactivity disorder. Design and descriptives. *European child & adolescent psychiatry* 24: 265-281. [crossref]

118. Younesi E, Hofmann-Apitius M (2013) Biomarker-guided translation of brain imaging into disease pathway models. *Scientific reports* 3. [crossref]

119. Davinelli S, Intrieri M, Russo C, Di Costanzo A, Zella D, et al. (2011) The "Alzheimer's disease signature": potential perspectives for novel biomarkers. *Immunity & Ageing* 8: 1. [crossref]

120. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nature methods* 8: 665-670. [crossref]

121. Bairoch AM, Apweiler R, Wu CH, Barker WC, Boeckmann B, et al. (2005) The universal protein resource (UniProt). *Nucleic acids research* 33: D154-9. [crossref]

122. Spasic I, Ananiadou S, McNaught J, Kumar A (2005) Text mining and ontologies in biomedicine: making sense of raw text. *Brief Bioinform* 6: 239-251. [crossref]

123. Hoehndorf R, Schofield PN, Gkoutos GV (2015) The role of ontologies in biological and biomedical research: a functional perspective. *Brief Bioinform* 16: 1069-1080. [crossref]

124. Köhler S, Bauer S, Mungall CJ, Carletti G, Smith CL, et al. (2011) Improving ontologies by automatic reasoning and evaluation of logical definitions. *BMC Bioinformatics* 12: 418. [crossref]

125. Smith B, Arabandi S, Brochhausen M, Calhoun M, Ciccarese P, et al. (2015) Biomedical imaging ontologies: A survey and proposal for future work. *J Pathol Inform* 6: 37. [crossref]

126. Buckler AJ, Ouellette M, Danagoulian J, Wernsing G, Liu TT, et al. (2013) Quantitative imaging biomarker ontology (QIBO) for knowledge representation of biomedical imaging biomarkers. *Journal of digital imaging* 26: 630-641. [crossref]

127. Lasbleiz J, Saint-Jalmes H, Duvauferrier R, Burgun A (2011) Creating a Magnetic resonance imaging ontology. *Studies in health technology and informatics* 169: 784. [crossref]

128. Lasbleiz J, Morelli J, Schnel N, Burgun A, Duvauferrier R, et al. (2012) MRI image artifact ontology: a proposed method for improved recognition. *Stud Health Technol Inform* 180: 103-107. [crossref]

129. Barillot C, Benali H, Dojat M, Gaignard A, Gibaud B, et al. (2006) Federating distributed and heterogeneous information sources in neuroimaging: the NeuroBase project. *In HealthGrid* 6: 3-13. [crossref]

130. Gibaud B, Kassel G, Dojat M, Batrancourt B, Michel F, et al. (2011) NeuroLOG: sharing neuroimaging data using an ontology-based federated approach. In AMIA... Annual Symposium proceedings [electronic resource]/AMIA Symposium. AMIA Symposium 2011: 472-480.

131. Iyappan et al., "Neuroimaging Feature Terminology (NIFT): a controlled terminology for the annotation of brain imaging features", Manuscript in final stages of submission

132. Gattermayer T (2007) SCAIView: annotation and visualization system for knowledge discovery (Doctoral dissertation, Masters thesis, Life Science Informatics at Bonn-Aachen International Center for Information Technology (B-IT), Germany).

133. Müller B, Fluck J, Hofmann-Apitius M, Friedrich CM BENCHMARKING SCAIView.

134. Snyder HM, Corriveau RA, Craft S, Faber JE, Greenberg SM, et al. (2015) Vascular contributions to cognitive impairment and dementia including Alzheimer's disease. *Alzheimer's & Dementia* 11: 710-717. [crossref]

135. Nizetic D, Chen CL, Hong W, Koo EH (2015) Inter-Dependent Mechanisms Behind Cognitive Dysfunction, Vascular Biology and Alzheimer's Dementia in Down Syndrome: Multi-Faceted Roles of APP. *Frontiers in behavioral neuroscience* 9. [crossref]

136. Aliev G, Obrenovich M E, Smith MA, Perry G (2003) Hypoperfusion, mitochondria failure, oxidative stress, and Alzheimer disease. *BioMed Research International* 2003: 162-163. [crossref]

137. Attems J, Jellinger KA (2014) The overlap between vascular disease and Alzheimer's disease--lessons from pathology. *BMC Med* 12: 206. [crossref]

138. O Akinyemi R, B Mukaetova-Ladinska E, Attems J, Ihara M, N Kalaria R (2013) Vascular risk factors and neurodegeneration in ageing related dementias: Alzheimer's disease and vascular dementia. *Current Alzheimer Research* 10: 642-653. [crossref]

139. Snyder HM, Corriveau RA, Craft S, Faber JE, Greenberg SM, et al. (2015) Vascular contributions to cognitive impairment and dementia including Alzheimer's disease. *Alzheimer's & Dementia* 11: 710-717. [crossref]

140. Iadecola C (2016) Vascular and Metabolic Factors in Alzheimer's Disease and Related Dementias: Introduction. *Cell Mol Neurobiol* 36: 151-154. [crossref]

141. Zhang N, Gordon ML, Goldberg TE (2016) Cerebral blood flow measured by arterial spin labeling MRI at resting state in normal aging and Alzheimer's disease. *Neuroscience & Biobehavioral Reviews* 72: 168-175. [crossref]

142. Valotassiou V, Angelidis G, Papatriantafyllou J, Sifakis N, Georgoulias P (2011) The Clinical Use of SPECT and PET Molecular Imaging in Alzheimer's Disease. INTECH Open Access Publisher. [crossref]

143. Bergmann H, Sinzinger H (2013) Radioactive Isotopes in Clinical Medicine and Research. Birkhäuser. [crossref]

144. Abe T, Usui A (2014) [Cerebral blood flow monitoring]. Kyobu geka. *The Japanese journal of thoracic surgery* 67: 623-629.

145. Promjunyakul NO, Lahna DL, Kaye JA, Dodge HH, Erten-Lyons D, et al. (2016) Comparison of cerebral blood flow and structural penumbras in relation to white

matter hyperintensities: A multi-modal magnetic resonance imaging study. *Journal of Cerebral Blood Flow & Metabolism* 39: 1528-1536. [crossref]

146. Wei X, Yan R, Chen Z, Weng R, Liu X, et al. (2016) Combined Diffusion Tensor Imaging and Arterial Spin Labeling as Markers of Early Parkinson's disease. *Scientific Reports* 6. [crossref]

147. Julien C, Tremblay C, Émond V, Lebbadi M, Salem Jr N, et al. (2009) SIRT1 decrease parallels the accumulation of tau in Alzheimer disease. *Journal of neuropathology and experimental neurology* 68: 48. [crossref]

148. Maiese K, Chong ZZ, Shang YC, Wang S (2011) Translating cell survival and cell longevity into treatment strategies with SIRT1. *Rom J Morphol Embryol* 52: 1173-1185. [crossref]

149. Chong ZZ, Shang YC, Wang S, Maiese K (2012) SIRT1: new avenues of discovery for disorders of oxidative stress. *Expert Opin Ther Targets* 16: 167-178. [crossref]

150. Kumar R, Chaterjee P, Sharma PK, Singh AK, Gupta A, et al. (2013) Sirtuin1: a promising serum protein marker for early detection of Alzheimer's disease. *PLoS One* 8: e61560. [crossref]

151. Hattori Y, Okamoto Y, Nagatsuka K, Takahashi R, Kalaria RN, et al. (2015) SIRT1 attenuates severe ischemic damage by preserving cerebral blood flow. *Neuroreport* 26: 113-117. [crossref]

152. Chen J, Zhou Y, Mueller-Steiner S, Chen LF, Kwon H, et al. (2005) SIRT1 protects against microglia-dependent amyloid-β toxicity through inhibiting NF-ϗB signaling. *Journal of Biological Chemistry* 280: 40364-40374. [crossref]

153. Tang BL, Chua CE (2008) SIRT1 and neuronal diseases. *Mol Aspects Med* 29: 187-200. [crossref]

154. Thakur N, Blanc PD, Julian LJ, Yelin EH, Katz PP, et al. (2010) COPD and cognitive impairment: the role of hypoxemia and oxygen therapy. *International journal of chronic obstructive pulmonary disease* 5: 263. [crossref]

155. Petegnief V, Planas AM (2013) SIRT1 regulation modulates stroke outcome. *Transl Stroke Res* 4: 663-671. [crossref]

156. Donmez G, Wang D, Cohen DE, Guarente L (2010) SIRT1 suppresses beta-amyloid production by activating the alpha-secretase gene ADAM10. *Cell* 142: 320-332. [crossref]

157. Ferguson D, Koo JW, Feng J, Heller E, Rabkin J, et al. (2013) Essential role of SIRT1 signaling in the nucleus accumbens in cocaine and morphine action. *J Neurosci* 33: 16088-16098. [crossref]

158. Herskovits AZ, Guarente L (2013) Sirtuin deacetylases in neurodegenerative diseases of aging. *Cell Res* 23: 746-758. [crossref]

159. Groot AJ, Vooijs MA (2012) The role of Adams in Notch signaling. In Notch Signaling in Embryology and Cancer Springer US 15: 36.

160. Kuhn PH, Wang H, Dislich B, Colombo A, Zeitschel U, et al. (2010) ADAM10 is the physiologically relevant, constitutive α-secretase of the amyloid precursor protein in primary neurons. *The EMBO journal* 29: 3020-3032. [crossref]

161. Suh J, Choi SH, Romano DM, Gannon MA, Lesinski AN, et al. (2013) ADAM10 missense mutations potentiate β-amyloid accumulation by impairing prodomain chaperone function. *Neuron* 80: 385-401. [crossref]

162. Kim M, Suh J, Romano D, Truong MH, Mullin K, et al. (2009) Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate α-secretase activity. *Human molecular genetics* 18: 3987-3996. [crossref]

163. Hosoi T, Nomura J, Ozawa K, Nishi A, Nomura Y (2015) Possible involvement of endoplasmic reticulum stress in the pathogenesis of Alzheimer's disease. *Endoplasmic Reticulum Stress in Diseases* 2. [crossref]

164. Giri M, Zhang M, Lü Y (2016) Genes associated with Alzheimer's disease: an overview and current status. *Clin Interv Aging* 11: 665-681. [crossref]

165. Darreh-Shori T, Forsberg A, Modiri N, Andreasen N, Blennow K, et al. (2011) Differential levels of apolipoprotein E and butyrylcholinesterase show strong association with pathological signs of Alzheimer's disease in the brain in vivo. *Neurobiology of aging* 32: 2320-e15. [crossref]

166. Ramanan VK, Risacher SL, Nho K, Kim S, Swaminathan S, et al. (2014) APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Mol Psychiatry* 19: 351-357. [crossref]

167. Vijayaraghavan S, Karami A, Aeinehband S, Behbahani H, Grandien A, et al. (2013) Regulated extracellular choline acetyltransferase activity—the plausible missing link of the distant action of acetylcholine in the cholinergic anti-inflammatory pathway. *PloS one* 8: e65936. [crossref]

168. Darreh-Shori T, Forsberg A, Modiri N, Andreasen N, Blennow K, et al. (2011) Differential levels of apolipoprotein E and butyrylcholinesterase show strong association with pathological signs of Alzheimer's disease in the brain in vivo. *Neurobiol Aging* 32: 2320. [crossref]

169. Ji H, Dai D, Wang Y, Jiang D, Zhou X, et al. (2015) Association of BDNF and BCHE with Alzheimer's disease: Meta-analysis based on 56 genetic case-control studies of 12,563 cases and 12,622 controls. *Experimental and therapeutic medicine* 9: 1831-1840. [crossref]

170. Reid GA, Darvesh S (2015) Butyrylcholinesterase-knockout reduces brain deposition of fibrillar β-amyloid in an Alzheimer mouse model. *Neuroscience* 298: 424-435. [crossref]

171. Manning EN, Barnes J, Cash DM, Bartlett JW, Leung KK, et al. (2014) APOEε4 is associated with disproportionate progressive hippocampal atrophy in AD. *PloS one* 9: e97608. [crossref]

172. Li B, Shi J, Gutman BA, Baxter LC, Thompson PM, et al. (2016) Influence of APOE genotype on hippocampal atrophy over time-An N= 1925 surface-based ADNI study. *PloS one* 11: e0152901. [crossref]

173. Cherbuin N, Sargent-Cox K, Easteal S, Sachdev P, Anstey KJ (2015) Hippocampal atrophy is associated with subjective memory decline: The PATH Through Life study. *Am J Geriatr Psychiatry* 23: 446-455. [crossref]

174. Parikh I, Fardo DW, Estus S (2014) Genetics of PICALM expression and Alzheimer's disease. *PLoS One* 9: e91242. [crossref]

175. Ramanathan A, Nelson AR, Sagare AP, Zlokovic BV (2015) Impaired vascular-mediated clearance of brain amyloid beta in Alzheimer's disease: the role, regulation and restoration of LRP1. *Frontiers in aging neuroscience* 7. [crossref]

176. Thomas RS, Henson A, Gerrish A, Jones L, Williams J, et al. (2016) Decreasing the expression of PICALM reduces endocytosis and the activity of β-secretase: implications for Alzheimer's disease. *BMC neuroscience* 17: 50. [crossref]

177. András IE, Toborek M (2013) Amyloid beta accumulation in HIV-1-infected brain: The role of the blood brain barrier. *IUBMB Life* 65: 43-49. [crossref]

178. Zlokovic BV, Deane R, Sagare AP, Bell RD, Winkler EA (2010) Low-density lipoprotein receptor-related protein-1: a serial clearance homeostatic mechanism controlling Alzheimer's amyloid β-peptide elimination from the brain. *Journal of neurochemistry* 115: 1077-1089. [crossref]

179. Sagare AP, Deane R, Zlokovic BV (2012). Low-density lipoprotein receptor-related protein 1: a physiological A? homeostatic mechanism with multiple therapeutic opportunities. *Pharmacology & therapeutics* 136: 94-105. [crossref]

180. Laifenfeld D, Patzek LJ, McPhie DL, Chen Y, Levites Y, et al. (2007) Rab5 mediates an amyloid precursor protein signaling pathway that leads to apoptosis. *The Journal of neuroscience* 27: 7141-7153. [crossref]

181. Zhao Z, Sagare AP, Ma Q, Halliday MR, Kong P, et al. (2015) Central role for PICALM in amyloid-Î² blood-brain barrier transcytosis and clearance. *Nat Neurosci* 18: 978-987. [crossref]

182. Sagare AP, Bell RD, Zlokovic BV (2013) Neurovascular defects and faulty amyloid-Î² vascular clearance in Alzheimer's disease. *J Alzheimers Dis* 33 Suppl 1: S87-100. [crossref]

183. Winkler EA, Nishida Y, Sagare AP, Rege SV, Bell RD, et al. (2015) GLUT1 reductions exacerbate Alzheimer's disease vasculo-neuronal dysfunction and degeneration. *Nature neuroscience* 18: 521-530. [crossref]

184. Lyros E, Bakogiannis C, Liu Y, Fassbender K (2014) Molecular links between endothelial dysfunction and neurodegeneration in Alzheimer's disease. *Current Alzheimer Research* 11: 18-26. [crossref]

185. Redzic Z (2011) Molecular biology of the blood-brain and the blood-cerebrospinal fluid barriers: similarities and differences. *Fluids and barriers of the CNS* 8: 1. [crossref]

186. De Giorgis V, Veggiotti P (2013) GLUT1 deficiency syndrome 2013: current state of the art. *Seizure* 22: 803-811. [crossref]

187. Shah K, DeSilva S, Abbruscato T (2012) The role of glucose transporters in brain disease: diabetes and Alzheimer's disease. *International journal of molecular sciences* 13: 12629-12655.

188. Duran-Aniotz C, Hetz C (2016) Glucose Metabolism: A Sweet Relief of Alzheimer's Disease. *Curr Biol* 26: R806-809. [crossref]

Copyright: ©2017 Iyappan A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

CONTRIBUTIONS FOR THE MANUSCRIPT

I am entirely responsible for performing all the analysis, data interpretation, and the writing of the manuscript: “Neuroimaging Feature Terminology: A Controlled Terminology for the Annotation of Brain Imaging Features.”

Dr. Erfan Younesi provided data for generating Table 1.

Alberto Redolfi provided raw patient image scans for manual annotation.

Shashank Khanna contributed to the compilation of Supplementary File 2.

Henri Vrooman contributed to the data for Supplementary File 3.

Journal of Alzheimer's Disease 59 (2017) 1153–1169
DOI 10.3233/JAD-161148
IOS Press

Neuroimaging Feature Terminology: A Controlled Terminology for the Annotation of Brain Imaging Features

Anandhi Iyappan[a,b], Erfan Younesi[a], Alberto Redolfi[c], Henri Vrooman[d], Shashank Khanna[a,b], Giovanni B. Frisoni[c,e] and Martin Hofmann-Apitius[a,b,*], for the Alzheimer's Disease Neuroimaging Initiative[1]

[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 Information Technology, Bonn, Germany*
[c]*Laboratory of Epidemiology and Neuroimaging, IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy*
[d]*Departments of Radiology and Medical Informatics, Biomedical Imaging Group Rotterdam, Erasmus MC University Medical Center, The Netherlands*
[e]*Memory Clinic and Laboratoire de Neuroimagerie du Vieillissement (LANVIE), University Hospitals and University of Geneva, Geneva, Switzerland*

Accepted 9 June 2017

Abstract. Ontologies and terminologies are used for interoperability of knowledge and data in a standard manner among interdisciplinary research groups. Existing imaging ontologies capture general aspects of the imaging domain as a whole such as methodological concepts or calibrations of imaging instruments. However, none of the existing ontologies covers the diagnostic features measured by imaging technologies in the context of neurodegenerative diseases. Therefore, the Neuro-Imaging Feature Terminology (NIFT) was developed to organize the knowledge domain of measured brain features in association with neurodegenerative diseases by imaging technologies. The purpose is to identify quantitative imaging biomarkers that can be extracted from multi-modal brain imaging data. This terminology attempts to cover measured features and parameters in brain scans relevant to disease progression. In this paper, we demonstrate the systematic retrieval of measured indices from literature and how the extracted knowledge can be further used for disease modeling that integrates neuroimaging features with molecular processes.

Keywords: Alzheimer's disease, annotation, brain, neuroimaging, terminology

INTRODUCTION

Brain imaging technologies have revolutionized the way that neurodegenerative diseases, such as Alzheimer's disease (AD), are diagnosed and tracked. Since the human brain is largely inaccessible for direct sampling, neuroimaging provides an alternative for measuring *in vivo* structural and functional features that can be used as biomarkers of disease onset and progression. The quantitative nature of imaging biomarkers, their potential to assess disease-modifying effects, and the ability to monitor the

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

*Correspondence to: Dr. Martin Hofmann-Apitius, Department of Bioinformatics, Fraunhofer Institute of Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, Sankt Augustin, 53754, Germany. 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).

safety of candidate drugs all make neuroimaging readouts an extensively used, measurable endpoint in clinical trials for neurodegenerative diseases [1]. Parameters and features that are clinically measured using neuroimaging biomarkers reflect biological or pathological changes underlying disease in the brain of patients; for example, positron emission tomography (PET) imaging measures the load of amyloid-β (Aβ) neuritic plaques through the uptake and binding of a particular radioligand in the living brain [2]. But such measurements at the clinical level are often disconnected from their underlying mechanistic causes, on one hand, and from their corresponding patient clinical tests, on the other hand.

The importance of neuroimaging in the new era of systems neurology is highlighted by its pivotal role in linking clinical readouts to underlying mechanistic changes [3]. Thus, there is a need for integrative approaches that enable multiscale modeling of both biological and clinical data with the aim of bridging the translational gap [4]. The first step toward this goal is, however, the collection and standardization of disparate and scattered data and knowledge across many resources available for research. Among several efforts in this direction, ApiNATOMY correlates brain imaging diagnostics to affected anatomical regions of the brain through the Foundational Model of Anatomy ontology (FMA) [5]; the OntoNeuroLOG ontology covers the domain of imaging datasets and their processing methods [6]; RadLex provides a lexicon of terms relevant to diagnostic and interventional radiology [7]; and the quantitative imaging biomarker ontology (QIBO) represents a series of heterogeneous concepts across several fields including imaging physics, contrast agents, biology, and quantitation techniques [8]. Neuroimaging Data Model and Taskforce (NIDM) facilitates the exchange of large publication corpus and other relevant metadata such as provenance information related to the neuroimaging research for establishing the reproducibility of research experiments as well as overcoming the challenge of data sharing (http://nidm.nidash.org/specs/nidm-overview.html).

In parallel to the generation of imaging datasets, an increasing amount of imaging information is published within literature articles which often report on measured features in patients with AD [9, 10]. Such studies typically try to correlate neuroimaging readouts with defined disease stages or subtypes. As an example, Whitwell and co-workers utilized magnetic resonance imaging (MRI) in patients with atypical variants of AD and were able to categorize these patient groups, based on measuring patterns of atrophy in medial temporal and cortical grey matter, into hippocampal sparing AD, limbic predominant AD, and typical AD subtypes [11]. This example clearly shows the importance of harvesting neuroimaging feature information from literature not only for monitoring critical imaging findings but also for stratification of patients based on their diagnostic status.

To this end, UMLS metathesaurus vocabularies were used to annotate and index radiology journal figure captions from more than 9000 articles for image information retrieval [12]. Similarly, RadLex was applied to the biomedical imaging literature and annotated more than 385,000 figures with RadLex terms [13]. However, when the National Cancer Institute Thesaurus (NCIT), Radiology Lexicon (RadLex; http://www.radlex.org/), Systemized Nomenclature of Medicine (SNOMED-CT), and International Classification of Diseases (ICD-9-CM) were evaluated for retrieval of radiology reports containing critical imaging findings, it was found that no single terminology is optimal for retrieving radiology reports with critical findings [14].

Biomedical terminologies and ontologies have proven their role in namespace harmonization and mediation of semantic interoperability in numerous examples [15]. One of the main application domains of shared semantics (ontologies and terminologies) lies in metadata annotation as well as data integration and knowledge retrieval [16]. The neuroimaging community has not yet come up with a consensus for commonly used and shared metadata. However, over the past decades, many initiatives have made their primary data publicly available [17]. Out of those, the Alzheimer Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu/) and Parkinson's Progression Markers Initiative (PPMI) (http://www.ppmi-info.org/) have gained increasing momentum for creating an impact on data sharing across the scientific community. Despite the ongoing efforts, the significant lack of structural and semantic interoperability impedes the momentum of data sharing [18] and lack of an established framework hampers the merging of imaging data from other resources [8]. The heterogeneous data such as cortical thickness or neuropsychological assessments of individual patients that are stored in ADNI/PPMI datasets do not follow a standard nomenclature, which makes them difficult to interpret or use for validation.

Motivated by the obvious need for a terminology that enables a systematic representation and retrieval

of features derived from neuroimaging techniques, we aimed at developing a Neuro-Imaging Feature Terminology (NIFT) to capture and organize the knowledge domain of structural and functional brain features as measured and represented by neuroimaging technologies in the domain of neurodegeneration. In this study, we demonstrate the value of NIFT for the identification and extraction of neuroimaging features in both Medline abstracts and full-text publications in the context of neurodegenerative disease pathology. We also demonstrate the applicability of NIFT for the annotation of imaging readouts in MRI and CT scans. Furthermore, we go beyond retrieval and annotation of imaging concepts by providing an example of how extracted neuroimaging features can be utilized for mechanistic modeling of disease pathology.

MATERIALS AND METHODS

The NIFT terminology is built based on a hierarchical knowledge representation system by organizing higher level concepts as root nodes followed by specific sub-classes organized under them; however, it is not an ontology as it uses simple hierarchical relationships but is capable of being leveraged to an ontology in the future. However, to be leveraged to an ontology, NIFT should undergo major changes in the current hierarchical structure based on ontology formalism definitions. NIFT in its current form provides a first substrate for the scientific community to elaborate its conceptual complexity and structure. The Protégé OWL editor was used to build this hierarchical terminology (http://protege.stanford.edu/). This terminology was constructed using the OWL language for two reasons: firstly, the hierarchical edition and annotation of concepts in OWL language facilitates creation of such a granular terminology: and secondly, the OWL format of NIFT ensures the interoperability of the terminology file. The concepts that are included under this terminology were examined by experts from the clinical research domain.

Generation of NIFT

The NIFT terminology concepts were gathered by collecting and reading relevant publications, e-books, websites, and medical blogs related to imaging in neurodegeneration. Following the initial literature search, we also adapted some concepts from already published, highly relevant ontologies such as QIBO [19] and Radlex. Ontologies such as QIBO and Radlex had well-structured concepts such as Imaging Techniques and Imaging Agents, which were contextually relevant for the development of NIFT. Essential entities used in the ADNI (http://adni.loni.usc.edu/) were also included in our terminology system.

Consequently, we enriched the NIFT with measured biomarkers obtained from the Biomedical Imaging Group Rotterdam (BIGR) pipeline, UMC Rotterdam [20] and neuGRID platform (http://www.neugrid4you.eu). The BIGR pipeline consists of six image processing pipelines such as FreeSurfer (http://freesurfer.net/), BIGR Tissue Segmentation [21], BIGR hippocampus segmentation [22], BIGR SAMSco [23], BIGR diffusion imaging pipeline [24], and Human Connectome Mapper [25]. The neuGRID platform consists of three image processing pipelines including: FreeSurfer, Adaboost, and SPMgrid. This platform was used to extract measured imaging indices to be added to the terminology.

For the sake of covering brain-specific anatomical structures in NIFT, we made use of the Brain Region & Cell Type Terminology (BRCT) which was initially developed to capture a wide range of key concepts representing human brain neuroanatomical structures and integrate their corresponding cell types (http://bioportal.bioontology.org/ontologies/BRCT). Alzheimer Disease Ontology [26] was also re-used to enrich NIFT. Pathway concepts were derived from the pathway Terminology System, that was developed with the intention to support the extraction of pathway information specific to the neurodegenerative disease domain [27].

Upon completion, this terminology system was reviewed by a clinical imaging expert team (Professor Frisoni's team at the University Hospitals Geneva) which further improved the quality and relevance of the classification.

Natural language processing (NLP)-based assessment of NIFT performance

In order to assess the relevancy of the NIFT terminology, we compared the performance of our terminology with the two already well-established imaging ontologies, QIBO and Biomedical Image Ontology (BIM) [28]. This comparison was performed at the level of terminologies found in those ontologies as they claim to capture the knowledge of image annotations and imaging biomarkers, respectively. To perform this, we used an NLP–based approach and ran the PDF tagger over the previously selected full-text publications (PMC1, PMC2,

PMC3, and PMC4) using all the three ontology/terminology systems and validated the retrieval of the maximum annotation of terms specific to neuroimaging domain. The validation of the terminology using PDF tagger was performed using the formula:

$$\text{mjk} = \sum_{i=1}^{nk} a\text{i}$$

where a, frequency of single term; j, document number; mj, overall frequency in document j; nk, number of items in dictionary k; k, dictionary number.

This index sums up the recall of relevant terms captured using the relevant terminology over all the terms found in the document. This sum gives an overall count of different concepts and terms captured from the given document.

This analysis was done to demonstrate the usability of NIFT in extracting relevant context from publications of interest.

Correlating clinical diagnosis with imaging features for staging AD

For bridging the clinical indices with imaging readouts, we systematically harvested relevant publications using the query "(((([Neuroimaging Feature]) AND [MeSH Disease "Alzheimer Disease"]) AND [Alzheimer Ontology: "Cognitive tests"]) AND [Organism: "Homo sapiens"]) AND [BRCT]" in SCAIView.

Retrieval and mining figure captions and full-text from PubMed

Following the curation and further refinement of the terminology, NIFT was integrated into our in-house literature mining environment SCAIView [29]. SCAIView enables the users to efficiently retrieve context specific articles from the literature using standardized terminologies and ontologies. NIFT in its SCAIView integrated form can be freely accessed using this link (http://academia.scaiview.com/academia//).

We performed an overall coverage analysis of NIFT by running it over figure captions and full-text articles using SCAIView. For this, we converted the OWL file into a dictionary (.syn) file using a java program. The resulting dictionary was incorporated in ProMiner, which is a rule-based entity recognition system [30]. The hierarchical structure of the OWL file was converted into an XML tree so that NIFT can be navigated within the SCAIView environment and faceted search becomes feasible. The ProMiner program was subsequently run over the five figure captions which were enriched with imaging indices from PubMed articles and four full-text publications from PubMed Central (PMC), which generated an abstract with markup of the terms specific for NIFT.

We also performed an analysis of full-text publications using a special PDF tagger (http://publica.fraunhofer.de/eprints/urn_nbn_de_0011-n-936860.pdf). In order to perform this task, we chose four full-text publications, which were relevant to the neurodegenerative context as well as reported imaging findings namely PMC1 [31], PMC2 [32], PMC3 [33], and PMC4 [34]. The PDF tagger was run over these publications for validation of the coverage of NIFT and results were stored in a dedicated directory. The PDF tagger first makes use of the documents in the directory as an input to create a term list from all the PDF files. Following the complete annotation of the PDF files, an output file was automatically generated with the original PDF file containing additional annotations highlighted through markup of terms.

Annotation of image scans using NIFT

In order to annotate brain scans with NIFT, we chose three groups of patients with different diagnostic features, namely: ADNI_016_S_4952 Control (CN), ADNI_002_S_4171 Mild Cognitive Impairment (MCI), and ADNI_003_S_4136 AD from the ADNI dataset (http://adni.loni.usc.edu/). ADNI is a large-scale, multicenter study which has been structured to develop molecular, clinical, and biochemical biomarkers from longitudinal patient data for early detection of AD. We processed PET (F18-AV-45 and FDG [18]) and T13D MPRAGE scans using the neuGRID platform with different pipelines, such as: SPMgrid to detect hypo-metabolism as well as amyloid burden; Freesurfer to highlight cortical thickness measurements and subcortical morphological differences; and Adaboost to quantify the hippocampal differences among the three diagnostic groups. The morphological changes observed from the patient scans were further annotated manually using the NIFT terminology.

Mechanistic modeling of image-derived indices in the context of AD

Yet another important aspect of this paper is to identify the role of molecular mechanisms, which bring about clear diagnostic outcomes captured by

imaging techniques. For this purpose, we generated the following query in SCAIView: "(([Neuroimaging Feature]) AND [MeSH Disease: "Alzheimer Disease"]) AND [Organism: "Homo sapiens"]" and filtered for Human Genes/Proteins. Next, we developed a "global map" of brain-region image-derived features along with molecular readouts, such as genes linked to a neuroimaging feature. We studied mechanisms of hippocampal atrophy in detail at molecular and cellular level. This knowledge was transformed into a cause-and-effect model using Biological Expression Language (BEL) platform (http://openbel.org/). BEL is a platform for representing causal and correlative relationships from biological context in a computer readable form. Then, we performed a high-resolution modelling of the mechanism underlying hippocampal atrophy. The outcome of this analysis will be further discussed in the Results section.

RESULTS

Often, literature resources misclassify an imaging technique as a biomarker while many others denote the derived indices as a biomarker. Owing to this, NIFT was constructed to represent, integrate, and harmonize heterogeneous knowledge across the domain of imaging biomarkers in the context of neurodegeneration.

Structure and content

NIFT comprises of 7 major classes namely Algorithms, Brain Region, Clinical indices, Clinical trial information, Imaging technique, Measured Feature, and finally Radiopharmaceutical compound. There are in total 1,221 terms in NIFT. The root concepts of NIFT include

(i) Algorithms which contains 4 children nodes namely: Image acquisition, MR-image analysis, PET-image analysis, and Post-processing algorithm. This concept contains all the brain imaging features that are automatically detected using various imaging pipelines such as FreeSurfer.

(ii) The second root concept in NIFT is Clinical Indices which has two children concepts namely AD and Parkinson's disease. This includes all the genetic, proteomic biomarkers mentioned in the literature for AD and Parkinson's disease.

(iii) The third root concept is Clinical trial information which contains three children concepts namely adverse effects observed in patients with neurodegeneration, neuropsychological assessments and scores such as Mini-Mental Status Examination score, Alzheimer's Disease Assessment Scale-Cog test, and clock draw test to name a few.

(iv) The fourth root concept of NIFT is Imaging Technique. This contains 7 children concepts, each of them represents the different imaging techniques used to study the various structural and functional dimensionality of the brain.

(v) The fifth root concept consists of measured features. This concept covers a wide range of "observable indicators" that determine the state of the brain and disease progression observed using various imaging techniques. This concept includes structural features such as cortical thickness, cerebral atrophy and functional features such as glucose metabolism, blood oxygenation level dependent signal.

(vi) The last root concept consists of radiopharmaceutical compounds. This concept contains all the radioactive tracers that are induced in the brain to diagnose dysfunction.

NIFT is available in OWL format and can be accessed from the following link (https://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/downloads.html). The hierarchical structure of NIFT is illustrated in Fig. 1.

NIFT evaluation

The content evaluation of NIFT in comparison to QIBO and BIM ontologies showed that NIFT performed comparatively better than QIBO and significantly better than BIM in capturing relevant terminology (see Fig. 2). For the first document (PMC1), we found 97 relevant terms annotated by BIM, 204 with NIFT, and 113 with QIBO. The second document (PMC2) was annotated with 308 relevant terms by BIM, 1334 terms by NIFT, and 1056 terms by QIBO. The third document (PMC3) retrieved 153 terms for BIM, 552 terms for NIFT, and 495 terms for QIBO. The fourth and final document (PMC4) retrieved 87 terms for BIM, 303 terms for NIFT, and 217 terms for QIBO. The PMC documents can be found in the Supplementary File 1.

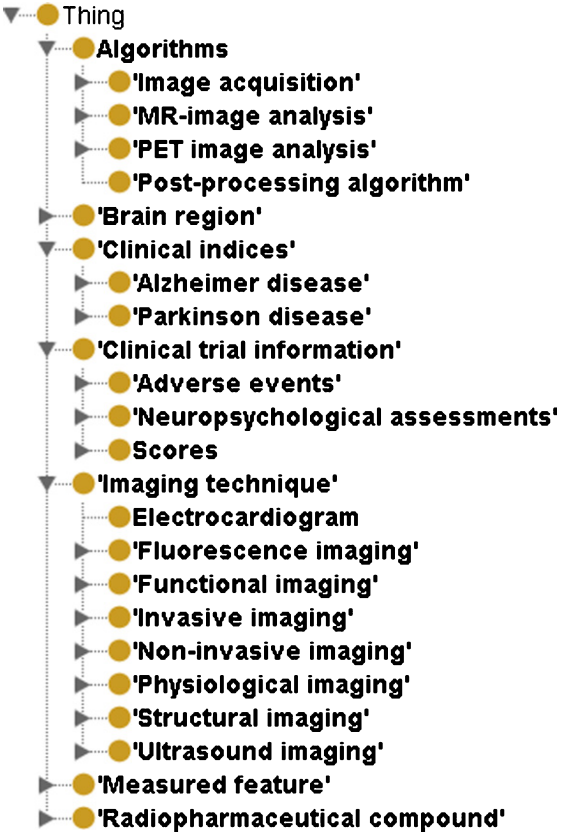


Fig. 1. Hierarchical structure of NIFT as visualized in the Protégé OWL Editor. This figure depicts the higher level concepts the terminology namely Algorithms, Brain Region, Clinical Indices, Clinical trial information, Imaging Technique, Measured Feature, and Radiopharmaceutical compound.

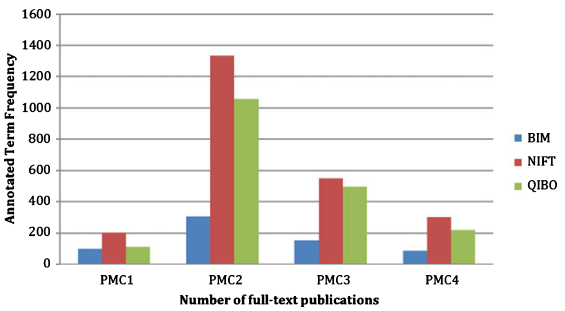


Fig. 2. Cross-validation of NIFT terminology against QIBO and BIM. The figure illustrates the evaluation of NIFT by comparing the term relevancy from NIFT, QIBO, and BIM against four full-text PubMed Central articles (PMC1, PMC2, PMC3, and PMC4).

The usability of well-annotated terminology systems can only be considered useful if they are applicable to relevant research. To assess the applicability of NIFT, we have studied the role of image derived indices for diagnosis and how they complement the clinical assessments for better disease prognosis. One of our aims was to establish a (plausible) bridge between clinical, imaging, and cognitive tests which is not only multi-modal, but should enable disease sub-type identification and classification. Our hope is that linking imaging with anatomical as well as diagnostic readouts in AD can help to gain better insight into disease progression and thereby provide more accurate diagnoses.

However, imaging-derived indices information is often scattered throughout the largely unstructured scientific literature, which needs to be analyzed in a systematic manner. Using NIFT, we retrieved 4,029 publications. Out of the 4,029 publications, we filtered 1000 documents that contained at least one quantitative imaging feature, one neuropsychological test at clinic, and a diagnosed stage of AD in corresponding patients (see the query in the Methods section). To exclude false positive documents, we manually curated all the 1000 publications and we found 101 articles that were relevant. For manual curation of documents, we followed a 3-step procedure which are as following:

(i) Only those articles that had informative relationship between neuropsychological assessment and radiological finding were considered for further analysis.
(ii) Articles that only had information about either neuropsychological assessment or radiological findings were not considered for further analysis
(iii) Articles that contained both neuropsychological assessment as well as radiological finding but did not have any meaningful relationship between them, were not considered for further analysis.

The resulting overview shows a pattern based on which imaging technologies and measured features derived from these technologies can be used to categorize the underlying clinical manifestations of patients and thereby links clinical and imaging readouts with the stage of the patient (see Table 1). The overall relation between the quantitative imaging feature, psychological feature, and diagnosis can be found in Supplementary File 2.

We also conducted a systematical analysis of the heterogeneous imaging techniques and readouts and combining them with the anatomical correlates and clinical endpoints. According to our analysis (as seen in Table 1), the medial temporal atrophy (MTA) as

Table 1
Mapping imaging derived features to clinical diagnosis and stage in Alzheimer's disease

PMID	IMAGING TECHNIQUE	IMAGING FEATURE	RELATION	NEUROPSYCHOLOGICAL ASSESSMENTS	BRAIN REGION	DIAGNOSIS
18468705	MRI	Medial Temporal Atrophy	Positive correlation	Mean visual rating scale (5.39)	Temporal lobe	Late AD
18468705	MRI	Medial Temporal Atrophy	Positive correlation	Mean visual rating scale (2.16)	Temporal lobe	Vascular dementia
18468705	MRI	Medial Temporal Atrophy	Positive correlation	Mean visual rating scale (0.56)	Temporal lobe	Healthy
11594917	MRI	Hippocampal atrophy	Correlation	Neuropsychological test ($r > 0.5$)	Hippocampus	Late AD
2061508	DTI	White matter integrity, brain atrophy	Associated	Clinical dementia rating (0.5)	Brain, white matter	Mild AD, EOAD
23880336	MRI	Cortical thinning	Associated	Montgomery Asberg Depression Rating Scale	Cortex	Mild AD, LBD
12834197	MRI	Hippocampal atrophy, corpus callosum atrophy	Associated	CDR (1 and 2)	Hippocampus Corpus callosum	Probable AD
12834197	MRI	Hippocampal atrophy	Associated	CDR (0.5)	Hippocampus	Mild AD
18400396	MRI	Cortical thinning	Associated	CDR (0.5)	Cortex	Pre-dementia

MRI, magnetic resonance imaging; DTI, diffusion tensor imaging; CDR, Clinical Dementia Rating; AD, Alzheimer's disease; EOAD, early onset Alzheimer's disease; LBD, Lewy body disease.

such is a common phenomenon observed in all the three diagnostic classes; however, they could be better distinguished as AD when the MTA score is the highest (5.39). Similarly, the atrophy could be classified as vascular dementia when the MTA score is 2.16 and in case of healthy patients, MTA could still occur, but with a very minimal score (0.56).

Analysis of literature for neuroimaging features

Mining image captions from literature

Although a fair amount of information on the image-derived findings is usually reported in the abstract of publications, specific features and interpretations gained from brain imaging experiments are often described in the caption of imaging figures that accompany the abstract text in PubMed. We, therefore, tested the relevance and performance of NIFT by applying it to a text mining scenario for analysis of figure legends extracted from publications. A typical example of figure captions annotated using NIFT terms that were extracted from PubMed abstracts is shown in Fig. 3. This figure highlights important quantitative biomarkers such as cortical ribbon which occurs due to the hyperintensity of the cortex observed in patients with early MCI and AD. This radiological sign can be detected using a diffusion tensor imaging technique and fractional anisotropy which is an important measure that demonstrates the connectivity of the brain as well as the tissue characteristics such as myelination and fiber density.

Mining full-text publications

In a separate experiment, we annotated a large corpus of full-text publications in order to examine the coverage of NIFT. A typical example for the automated annotation of a section of a full-text publication is shown in Fig. 4. This figure highlights the coverage of the NIFT terms from the full-text publication which includes neuropsychological assessments, brain regions, imaging technique as well as imaging biomarkers. This application demonstrates the usability of NIFT in mining context-specific, full-text publications in the field of neurodegeneration. Retrieval of context-specific, full-text publications can further be used to build a gold-standard corpus in the neurodegeneration domain for generation of novel hypotheses.

Annotation of image scans using NIFT

In a separate experiment, we tested, to what extend NIFT terms are suitable for the annotation of

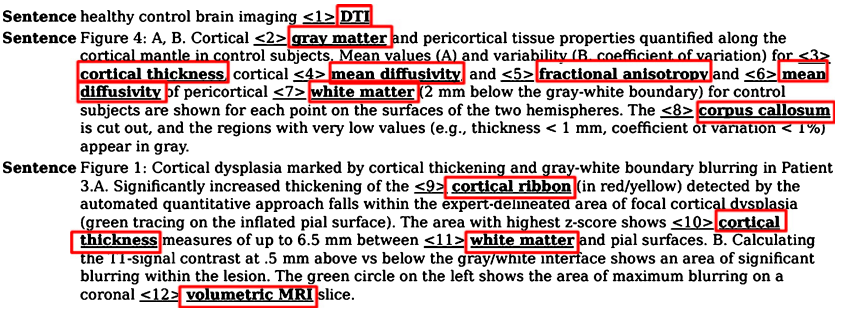


Fig. 3. Annotation of an assembly of figure captions with NIFT terminology. This figure showcases the figure captions extracted from publications using NIFT terminology. The red box indicates the NIFT terms present in the figure captions.

atrophy scores (Korf et al. 2004). More recently, automated techniques to extract volumes of interest and cortical thickness values for numerous neocortical regions (Dale et al. 1999, Fischl & Dale 2000), as well as semiautomated whole-brain morphometry techniques, such as voxel-based morphometry (Ashburner & Friston 2000) and other techniques (Fan et al. 2008, Hua et al. 2008) that determine the density or volume of gray matter (GM), white matter (WM), and CSF on a voxel-by-voxel basis, have been developed and utilized in studies of brain aging and AD. More advanced structural imaging techniques, including diffusion weighted imaging and diffusion tensor imaging (DWI/DTI), magnetic resonance spectroscopy (MRS), and perfusion imaging have also been used to evaluate changes in patients with AD and those in prodromal stages. DWI/DTI techniques measure the integrity of WM tracts using two types of measures: (*a*) fractional anisotropy (FA), which reflects the diffusion direction of water in the fiber tracts and is thought to be a general measure of axonal integrity; and (*b*) mean diffusivity (MD) or apparent diffusion coefficient (ADC), which measures the overall diffusivity. Reduced FA and increased MD/ADC are considered to be markers of neuronal fiber loss and WM atrophy. MRS is a noninvasive technique allowing the measurement of biological metabolites in target tissue that has been used in studies of brain aging and AD. Two major metabolites that consistently show alterations in patients with AD and MCI include N-acetylaspartate (NAA) and myo-inositol. Finally, two MRI methods have traditionally been employed to measure cerebral perfusion in studies of AD, including dynamic susceptibility contrast enhanced MRI, which involves the injection of contrast agents to measure CBF and regional cerebral blood volume, and a relatively new technique called arterial spin labeling, which measures CBF without any external contrast agents by using magnetic pulses that label blood entering the brain (Alsop et al. 2010, Wang et al. 2011).

MRI can also be used to measure brain function. Functional MRI (fMRI) measures brain activity during a cognitive, sensory, or motor task or at rest by measuring blood flow and blood oxygen levels. The primary outcome measured in fMRI studies is a blood-oxygenation-level-dependent (BOLD) signal in which regional brain activity is assessed via changes in local blood flow and oxygenation (Ogawa et al. 1992). Given that activity-related brain metabolism is tightly coupled to regional blood oxygenation and flow (i.e., blood flow increases to keep the regional blood oxygen level high during brain activation and associated increases in metabolic demand), the BOLD signal is a useful measure for brain activation (Logothetis et al. 2001). However, altered coupling of neuronal metabolism and blood flow due to brain atrophy and/or hypoperfusion may cause alterations in the BOLD signal. Therefore, studies in older patient populations with brain atrophy and hypoperfusion, such as MCI and AD, should be evaluated and interpreted with these considerations in mind.

Fig. 4. Annotation of a section of a full-text article using the NIFT terminology. The ProMiner tagger was used to identify NIFT terms in full text; matching terms are marked up in red.

primary neuroimaging data (brain scans). Figure 5 depicts the comparison between control, MCI, and AD patient brain scans based on: the amyloid burden through [18] AV45-PET, the regional pattern of hypometabolism through FDG-PET, and hippocampal volumetry as well as cortical thickness through T13D MP-RAGE brain scans. The top part of the figure shows the amyloid burden and the hypometabolic clusters across the different brain regions. As it can be seen in Fig. 5, the control does not have any

amyloid deposit cluster and no hypometabolism detected, while in MCI, hypometabolic pattern starts to appear in the left hemisphere and more extensively in AD. The expected hypometabolic topography spread across the temporo-parietal regions, precuneus, and posterior cingulate cortex. All the patient-derived image scans can be found in Supplementary File 3.

Mechanistic modeling of imaging features in the context of AD pathology

Generating links between molecular entities and imaging modalities, even if very demanding and complex, could provide interesting insights into the disease progression as well as help to raise our understanding of the underlying pathology. On that note, we tried to establish that link by querying our in-house SCAIView tool for genes/proteins relevant to imaging features (See Methods section). We retrieved 1,853 gene/protein entities, out of which we identified the top 20 entities confined to interesting brain regions such as cortex, hippocampus, temporal lobe,and cerebrum. Using these entities, we produced the 'global map' of genes and imaging features (See Fig. 6). We also inferred from this model that these top ranking genes play a role in cortical thickness, hippocampal atrophy, temporal lobe atrophy, grey matter atrophy and cerebral atrophy, as follows.

Cortical thickness

Our systematic analysis of the literature revealed that many key players contributed to thinning of the cortex, which is a strong indicator of AD progression. In the following, we demonstrate lines of evidence about factors causally involved in or correlated with cortical thinning and exemplify their corresponding BEL codes:

- Increased expression of APP results in the accumulation of Aβ, which affects the thinning of the cortex [35, 36].

p(HGNC:APP) -> a("Amyloid beta-Peptides")
a("Amyloid beta-Peptides") – a(NIFT: "Cortical thinning")
a(NIFT: "Cortical thinning") -> path(MESHD: "Alzheimer Disease")

- Increased expression of CHI3L1, a gene responsible for inflammatory response [37, 38], was found to be correlated with cortical thickness [39].

p(HGNC:CHI3L1) -> bp(GO:"inflammatory response")
bp(GO:"inflammatory response") negativeCorrelation a(NIFT: "Cortical thickness")

- PSEN1 was found to cause neuronal loss [40, 41], which results in the shrinkage of the cortex due to neuronal injury [42, 43].

p(HGNC:PSEN1) -> bp (GO:"neuronal loss")
bp(GO:"neuronal loss") -> a(NIFT: "Cortical thinning")

- Well-known genes such as APOE4 along with APOE4 and BCHE carriers contributed to the structural alteration of the cortex, resulting in cortical thinning [44–47].

p(HGNC:APOE) -> a(NIFT: "Cortical thinning")
p(HGNC:BCHE) -> a(NIFT: "Cortical thinning")

- Some genes can be linked to cortical thinning through genetics approaches: genes such as FJ10357 [48], TOMM40 [49], and BDNF [50, 51] play a protective role in preserving the structure of the cortex, however, their genetic alteration results in cortical thinning–rs3748348, rs10524523 and rs6265, respectively.

p(HGNC: FJ10357) -> a(NIFT: "Cortical thickness")
g(dbSNP: rs3748348) – a(NIFT: "Cortical thinning")
p(HGNC: TOMM40) -| a(NIFT: "Cortical thickness")
g(dbSNP: rs10524523) – a(NIFT: "Cortical thinning")
p(HGNC: BDNF) -> a(NIFT: "Cortical thickness")
g(dbSNP: rs6265) – a(NIFT: "Cortical thinning")

Temporal lobe atrophy

We investigated two genes, APOE ε4 and TREM2, which mainly contribute to the atrophy of temporal lobes. TREM2 is an inflammatory response gene predominantly found in microglia [52, 53]. They are known to enhance phagocytosis as well

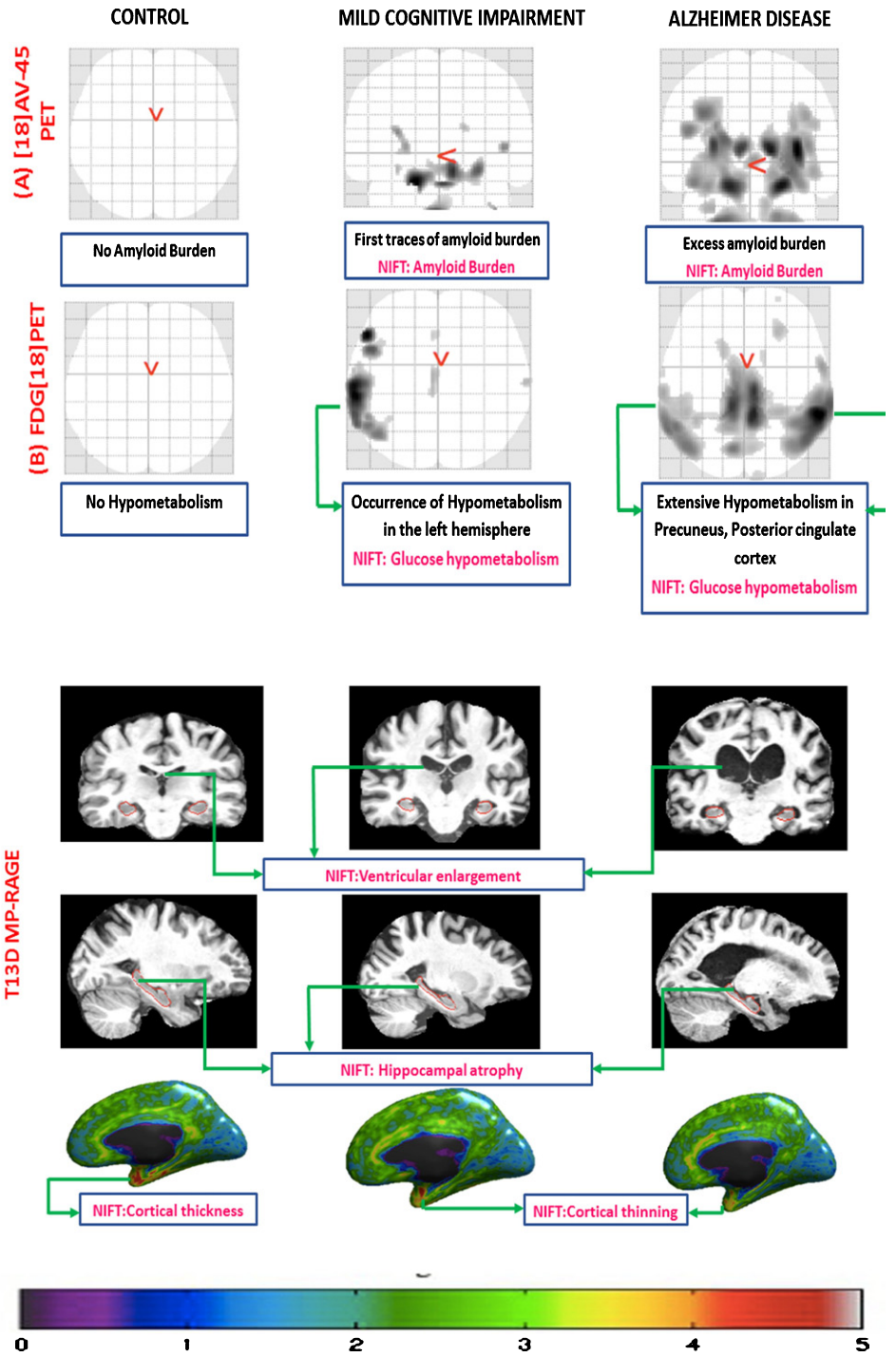


Fig. 5. Manual annotation of brain image scans using NIFT. This figure represents different biomarkers captured using three different imaging techniques in control, mild cognitive impairment (MCI), and AD respectively. A) [18] AV-45 PET scan: this figure captures the increased amount of amyloid burden (p-value threshold 0.001; voxel extend 10; smoothing kernel [8-8-8]) during the disease progression across CN, MCI, and AD, respectively. B) FDG [18] PET: this figure captures no hypometabolism in control, increased hypometabolic pattern in case of MCI, and extensive hypometabolic topography in the temporo-parietal regions, precuneus, and posterior cingulate cortex (p-value threshold 0.001; voxel extend 10; smoothing kernel [8-8-8]). C) T13D MP-RAGE: the first row of the figure demonstrates the progressive ventricular enlargement among control, MCI, and AD respectively. The second row represents progressive hippocampal atrophy across control, MCI, and AD. The third row represents progressive cortical shrinkage in the temporal-parietal lobe, posterior cingulate and precuneus area.

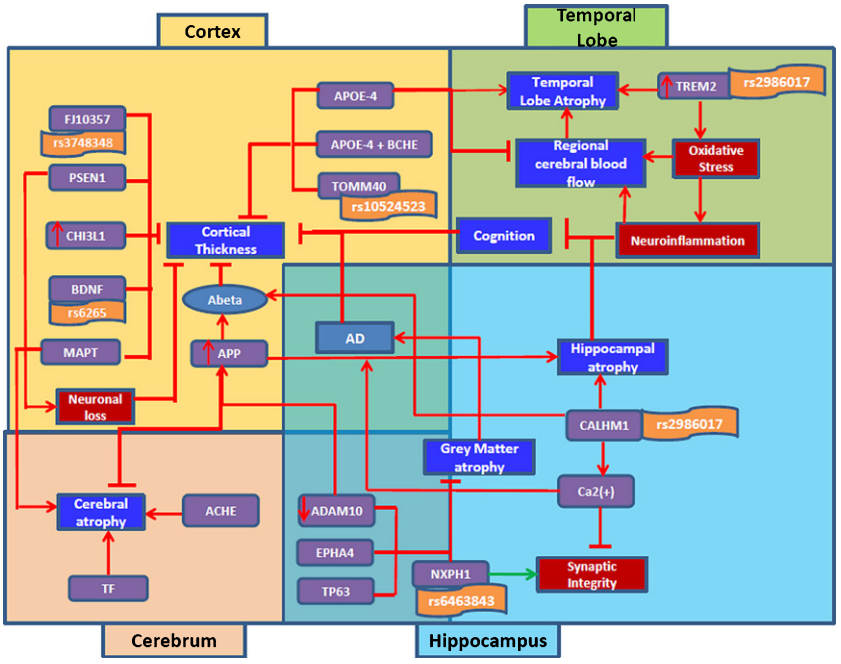


Fig. 6. Integrative view of literature-derived associations between molecular and clinical indices in AD through image-derived features. This figure illustrates the complex interaction of genetic players playing a causative/protective role in underlying disease pathology through neuroimaging indices. Top left part of the figure key genetic factors that play a role in shrinking of the cortex eventually leading to AD; top right part of the figure consists of genes involved in neuro-inflammation and temporal lobe atrophy; Bottom left part of the figure displays genes involved in cerebral atrophy; bottom right part consists of genes playing a role in hippocampal and gray matter atrophy. The red color symbol (-|) indicates perturbation of a gene. The red color arrow indicates the function of a gene in disease condition. The green arrow represents the normal process.

as maintaining cytokine production so that inflammatory responses can be triggered by TREM-1, a novel receptor expressed on neutrophils and monocytes [54, 55]. However, the genetic mutation of TREM2, rs75932628, causes the atrophy of temporal lobes through enhancing oxidative stress [56], which in turn causes the reduction of cerebral blood flow leading to reduction in regular supply of oxygen, glucose, and other nutrients to the temporal lobe, and finally the shrinkage of the temporal lobe [57, 58]. On the other hand, increased expression of APOE ε4 allele affects the flow of cerebral blood, further contributing to atrophy [59, 60].

Hippocampal atrophy

Through our work, we identified an interesting gene, CALHM1 which was known to regulate Aβ clearance [61] through the activation of insulin-degrading enzyme [62]. However, a genetic mutation by rs2986017 results in (i) loss of hippocampal neurons further causing atrophy as well as (ii) increased Aβ levels and altered calcium homeostasis which could result in reduced synaptic integrity and mitochondrial dysfunction [63].

Grey matter atrophy

Here, we identified a gene namely NXPH1, which was found to play a role in adhesion of dendrites and axons and maintaining synaptic integrity. However, the mutation of the gene, rs6463843, affects the synaptic integrity and results in loss of grey matter density leading to atrophy [64]. Apart from that, EPHA4 was also found to play a protective role in the glial glutamate transport that ultimately regulates hippocampal function as well as the maintenance of grey matter density [65, 66].

Cerebral atrophy

Cerebral atrophy was found to be regulated by two key players—Transferrin as well as ACHE. Transferrin was found to play a significant role in iron homeostasis [67]. However, the alteration of the gene could result in iron overload which causes damage to the cerebral structure, ultimately leading to cerebral

atrophy [68, 69]. On the other hand, the altered function of ACHE in the cholinergic system could result in the loss of cerebral neurons leading to cerebral atrophy [70]. The computer readable BEL version of this model is found in Supplementary File 4.

DISCUSSION

A "biomarker" is an accurately measured medical sign that indicates the medical state of the patient. However, in the field of imaging, this term is often misinterpreted due to the lack of standardization of terminology and concepts. Often, literature resources misclassify an imaging technique as a biomarker while many others denote the derived indices as a biomarker [71, 72]. Owing to this, NIFT was constructed to represent, integrate, and harmonize heterogeneous knowledge across the domain of imaging biomarkers in the context of neurodegeneration. NIFT serves as a single resource of the standard terminology that describes the domain of neuroimaging biomarkers in a hierarchical manner and has been designed to capture relevant image-derived features ("indices") with high specificity and granularity. As shown by our analysis, what distinguishes NIFT from other existing resources is the inclusion of various concepts ranging from algorithms that automate the process of measuring features to radiological tracers that help in revealing functional alterations of the brain. Such a standard reference terminology has the potential not only to support organization and exchange of imaging information among neurologists and clinical researchers but also to provide a useful tool for annotation of brain scan metadata as detection of meta-information in brain scans helps inferring neuroanatomical relationships present in imaging data [73]. With such an inventory, it is indeed possible to automatically extend the annotation of scans by incorporating NIFT in image annotation tools. Since NIFT combines specificity and granularity of imaging features in the context of neurology knowledge domain, users can intuitively navigate through different levels of concept granularity within a search engine and for instance, perform faceted search in literature mining environments.

With respect to the contextual specificity, as benchmark analysis of NIFT against two other highly domain-specific, relevant terminologies showed the overall granularity of medical relevant terms and cognitive tests in NIFT was comparably high, making NIFT a reference terminology resource specific to neuroimaging. The applicability of NIFT could be extended toward information retrieval and extraction. As demonstrated earlier, using NIFT for literature mining improves retrieval of the relevant, informative neuroimaging publications and supports curation and extraction of captured information from unstructured text. In the presence of other terminology sources, powerful filtering for faceted searches can be implemented. For instance, we can combine NIFT with HypothesisFinder [74] to systematically harvest speculative statements linked to imaging features; or combination of NIFT terms with ADO terms will allow us to systematically harvest factual statements that link imaging readouts to aspects of AD progression in literature; and finally, we also have the possibility of mining "shared imaging features" amongst other diseases by making use of the already integrated Parkinson Disease Ontology [75] and Multiple Sclerosis Ontology [76]. This could lead to domain specific imaging feature identification across disease scales.

Importantly, the usage of NIFT is not limited to information retrieval and extraction. Since the major mission of the neuroscience community currently is to bridge the gap between molecular mechanisms and imaging readouts, NIFT can be used to address this challenge by bringing context to computational modeling efforts. To demonstrate this possibility, we showed how NIFT serves as a valuable resource to support mechanistic modeling of complex AD pathomechanisms. As highlighted in Fig. 6, this high resolution mechanistic model captures novel genetic players such as CALHM1, NXPH1, and ADAM10, which cause hippocampal atrophy through neuronal loss. Here, we identified the various roles played by CALHM1 in AD pathology, ranging from controlling cytosolic Ca(2+) concentrations and Aβ levels to increased oxidative stress through glutamatergic neurotransmission inhibition [77, 78]. Similarly, another two novel genetic biomarkers were CHI3L1 and CAND1. CHI3L1, a protein that encodes YKL-40, was found to be associated with cortical thinning and was found to play a role in neuroinflammatory response. They were found to play a role in cell morphology and behavior; however, their association with susceptibility to AD has only been recently studied [79].

The current neuropathological studies on AD suggest that the clinical onset of the disease goes decades before the formation of neurofibrillary tangles and Aβ plaques [80, 81]. This brings up the need for heterogeneous measurable indicators that can aid

systematic tracing of alternative patterns of disease progression. ADNI have positioned themselves as pioneers in assembling patient records with cognitive and longitudinal assessments along with genetic and fluid sample measures. This interesting combination of measured metadata could provide unique insights into measurable signs before the expected onset of the disease. However, the challenge still remains to identify those patterns at an earlier stage through the use of combinatorial features. Furthermore, we foresee the option to perform systematic association studies in the literature between SNPs and mutations on one side and imaging features on the other side (Iyappan et al., in preparation). The multi-level association between genetic factors and clinical readouts can be directly used for modeling and mining across scales in the neurology and psychiatry field. A first attempt of demonstrating such systematic harvesting approach is the association of imaging features with cognition readouts (refer to Table 1). Such associations lead to comprehensive analysis of imaging features correlating with cognition.

To the best state of our knowledge, NIFT is the first reference compendium, which apprehends the various aspects of the derived quantitative measures from neuroimaging scans. We invite the scientific community to contribute to edition and enrichment of NIFT so that it can be leveraged to the level of a formal ontology in future.

Conclusion

To our knowledge, there have been little efforts invested so far in the direction of standardizing and capturing observable clinical imaging features, particularly in the neurology domain. Through this work, we attempted to bridge "omics" and imaging/clinical level data. This type of integration across scales is often regarded as the "holy grail" of integrative modeling and mining. Future approaches should be able to represent and model the disease progression in a longitudinal model by integrating molecular processes and imaging features over time, provided that longitudinal data capture would be extended to other omics data types beside imaging. For this purpose, we obviously need trajectories. Currently, the BEL modeling framework does not deliver this time dimension. However, we are working towards the extension of BEL by a time dimension. A long term perspective of this extension is the vision of a virtual patient cohort that comprises several such longitudinal "trajectories" representing the dynamics of important imaging features. The link between imaging and genetics will be a cornerstone for the construction of the virtual cohort; the generation of a "virtual dementia cohort" has recently been made a task in IMI-project AETIONOMY and we will see first results of the simulation of entire trials based on a "virtual dementia cohort" in the near future. The imaging derived features captured through NIFT will have a major role in that "virtual dementia cohort". We believe it would be desirable to generate a "metadata atlas" of the brain populated with NIFT concepts. Such an atlas could serve as a template for qualitative models that integrate imaging features from different, heterogeneous studies.

ACKNOWLEDGMENTS

The authors wish to acknowledge Frisoni's group for curation; Bernard de Bono for constructive discussions and advice; and students from the Life Science Informatics curriculum at Bonn-Aachen International Center for Information Technology (B-IT) for their support with curation and annotation tasks.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the

National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under AETIONOMY grant agreement n°115568, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/16-1148r2).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-161148.

REFERENCES

[1] Cash DM, Rohrer JD, Ryan NS, Ourselin S, Fox NC (2014) Imaging endpoints for clinical trials in Alzheimer's disease. *Alzheimers Res Ther* **6**, 87.

[2] Ishii K (2014) PET approaches for diagnosis of dementia. *AJNR Am J Neuroradiol* **35**, 2030-2038.

[3] Younesi E, Hofmann-Apitius M (2013) Biomarker-guided translation of brain imaging into disease pathway models. *Sci Rep* **3**, 3375.

[4] Kassab GS, An G, Sander EA, Miga MI, Guccione JM, Ji S, Vodovotz Y (2016) Augmenting surgery via multi-scale modeling and translational systems biology in the era of precision medicine: A multidisciplinary perspective. *Ann Biomed Eng* **44**, 2611-2625.

[5] De Bono B, Safaei S, Grenon P, Nickerson DP, Alexander S, Helvensteijn M, Kok JN, Kokash N, Wu A, Yu T, Hunter P, Baldock RA (2015) The open physiology workflow: Modeling processes over physiology circuitboards of interoperable tissue units. *Front Physiol* **6**, 24.

[6] Gibaud B, Kassel G, Dojat M, Batrancourt B, Michel F, Gaignard A, Montagnat J (2011) NeuroLOG: Sharing neuroimaging data using an ontology-based federated approach. *AMIA Annu Symp Proc* **2011**, 472-480.

[7] Langlotz CP (2006) RadLex: A new method for indexing online educational materials. *Radiographics* **26**, 1595-1597.

[8] Buckler AJ, Ouellette M, Danagoulian J, Wernsing G, Liu TT, Savig E, Suzek BE, Rubin DL, Paik D (2013) Quantitative imaging biomarker ontology (QIBO) for knowledge representation of biomedical imaging biomarkers. *J Digit Imaging* **26**, 630-641.

[9] Soucy J-P, Bartha R, Bocti C, Borrie M, Burhan AM, Laforce R, Rosa-Neto P (2013) Clinical applications of neuroimaging in patients with Alzheimer's disease: A review from the Fourth Canadian Consensus Conference on the Diagnosis and Treatment of Dementia 2012. *Alzheimers Res Ther* **5**, S3.

[10] Gispert JD, Rami L, Sánchez-Benavides G, Falcon C, Tucholka A, Rojas S, Molinuevo JL (2015) Nonlinear cerebral atrophy patterns across the Alzheimer's disease continuum: Impact of APOE4 genotype. *Neurobiol Aging* **36**, 2687-2701.

[11] Whitwell JL, Jack CR, Knopman DS, Boeve BF, Petersen RC, Josephs KA, Parisi JE, Dickson DW, Murray ME, Whitwell JL, Dickson DW, Murray ME, Weigand SD, Tosakulwong N, Senjem ML, Knopman DS, Boeve BF, Parisi JE, Petersen RC, Ord C, Jack R, Josephs KA (2012) Neuroimaging correlates of pathologically defi ned subtypes of Alzheimer's disease: A case-control study. *Lancet Neurol* **11**, 868-877.

[12] Kahn CE, Rubin DL (2009) Automated semantic indexing of figure captions to improve radiology image retrieval. *J Am Med Inform Assoc* **16**, 380-386.

[13] Kahn CE (2014) Annotation of figures from the biomedical imaging literature: A comparative analysis of radlex and other standardized vocabularies. *Acad Radiol* **21**, 384-392.

[14] Warden GI, Lacson R, Khorasani R (2011) Leveraging terminologies for retrieval of radiology reports with critical imaging findings. *AMIA Annu Symp Proc* **2011**, 1481-2588.

[15] Tao C, Pathak J, Solbrig HR, Wei W-Q, Chute CG (2013) Terminology representation guidelines for biomedical ontologies in the semantic web notations. *J Biomed Inform* **46**, 128-138.

[16] Hofmann-Apitius M (2015) Is dementia research ready for big data approaches? *BMC Med* **13** 145.

[17] Poldrack RA, Gorgolewski KJ (2014) Making big data open: Data sharing in neuroimaging. *Nat Neurosci* **17**, 1510-1517.

[18] Breeze JL, Poline J-B, Kennedy DN (2012) Data sharing and publishing in the field of neuroimaging. *Gigascience* **1**, 9.

[19] Buckler AJ, Ouellette M, Danagoulian J, Wernsing G, Liu TT, Savig E, Suzek BE, Rubin DL, Paik D (2013) Quantitative imaging biomarker ontology (QIBO) for knowledge representation of biomedical imaging biomarkers. *J Digit Imaging* **26**, 630-641.

[20] Vrooman HA, Cocosco CA, van der Lijn F, Stokking R, Ikram MA, Vernooij MW, Breteler MMB, Niessen WJ (2007) Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage* **37**, 71-81.

[21] Fortunati V, Verhaart RF, van der Lijn F, Niessen WJ, Veenland JF, Paulides MM, van Walsum T (2013) Tissue segmentation of head and neck CT images for treatment planning: A multiatlas approach combined with intensity modeling. *Med Phys* **40**, 71905.

[22] van der Lijn F, den Heijer T, Breteler MMB, Niessen WJ (2008) Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *Neuroimage* **43**, 708-720.

[23] de Boer R, Schaap M, van der Lijn F, Vrooman HA, de Groot M, van der Lugt A, Ikram MA, Vernooij MW, Breteler MMB, Niessen WJ (2011) Statistical analysis of minimum cost path based structural brain connectivity. *Neuroimage* **55**, 557-565.

[24] Guyader JM, Bernardin L, Douglas NH, Poot DH, Niessen WJ, Klein S (2015) Influence of image registration on apparent diffusion coefficient images computed from free-breathing diffusion MR images of the abdomen. *J Magn Reson Imaging* **42**, 315-330.

[25] Daducci A, Gerhard S, Griffa A, Lemkaddem A, Cammoun L, Gigandet X, Meuli R, Hagmann P, Thiran JP (2012) The Connectome Mapper: An open-source processing pipeline to map connectomes with MRI. *PLoS One* **7**, e48121.

[26] Malhotra A, Younesi E, Gündel M, Müller B, Heneka MT, Hofmann-Apitius M (2014) ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease. *Alzheimers Dement* **10**, 238-246.

[27] Iyappan A, Gündel M, Shahid M, Wang J, Li H, Mevissen HT, Müller B, Fluck J, Jirsa V, Domide L, Younesi E, Hofmann-Apitius M (2016) Towards a pathway inventory of the human brain for modeling disease mechanisms underlying neurodegeneration. *J Alzheimers Dis* **52**, 1343-1360.

[28] Bukhari AC, Nagy ML, Krauthammer M, Ciccarese P, Baker CJO (2015) BIM: An open ontology for the annotation of biomedical images. *CEUR Workshop Proc* **1515**, 1-5.

[29] Hofmann-Apitius M, Fluck J, Furlong L, Fornes O, Kolárik C, Hanser S, Boeker M, Schulz S, Sanz F, Klinger R, Mevissen T, Gattermayer T, Oliva B, Friedrich CM (2008) Knowledge environments representing molecular entities for the virtual physiological human. *Philos Trans A Math Phys Eng Sci* **366**, 3091-3110.

[30] Hanisch D, Fundel K, Mevissen H-T, Zimmer R, Fluck J (2005) ProMiner: Rule-based protein and gene entity recognition. *BMC Bioinformatics* **6**(Suppl 1), S14.

[31] Trojanowski JQ1, Vandeerstichele H, Korecka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter WZ, Weiner MW, Jack CR Jr, Jagust W, Toga AW, Lee VM, Shaw LM; Alzheimer's Disease Neuroimaging Initiative (2010) Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects. *Alzheimers Dement* **6**, 230-238.

[32] Risacher SL, Saykin AJ (2013) Neuroimaging and other biomarkers for Alzheimer's disease: The changing landscape of early detection. *Annu Rev Clin Psychol* **9**, 621-648.

[33] Risacher SL, Saykin AJ (2013) Neuroimaging biomarkers of neurodegenerative diseases and dementia. *Semin Neurol* **33**, 386-416.

[34] Shokouhi S, Claassen D, Riddle W (2014) Imaging brain metabolism and pathology in Alzheimer's disease with positron emission tomography. *J Alzheimers Dis Parkinsonism* **4**, pii: 143.

[35] Becker JA, Hedden T, Carmasin J, Maye J, Rentz DM, Putcha D, Fischl B, Greve DN, Marshall GA, Salloway S, Marks D, Buckner RL, Sperling RA, Johnson KA (2011) Amyloid-β associated cortical thinning in clinically normal elderly. *Ann Neurol* **69**, 1032-1042.

[36] Kim CM, Hwang J, Lee JM, Roh JH, Lee JH, Koh JY; Alzheimer's Disease Neuroimaging Initiative (ADNI) (2015) Amyloid beta-weighted cortical thickness: A new imaging biomarker in Alzheimer's disease. *Curr Alzheimer Res* **12**, 563-571.

[37] Bhardwaj R, Yester JW, Singh SK, Biswas DD, Surace MJ, Waters MR, Hauser KF, Yao Z, Boyce BF, Kordula T (2015) RelB/p50 complexes regulate cytokine-induced YKL-40 expression. *J Immunol* **194**, 2862-2870.

[38] Melah KE, Lu SY-F, Hoscheidt SM, Alexander AL, Adluru N, Destiche DJ, Carlsson CM, Zetterberg H, Blennow K, Okonkwo OC, Gleason CE, Dowling NM, Bratzke LC, Rowley HA, Sager MA, Asthana S, Johnson SC, Bendlin BB (2016) Cerebrospinal fluid markers of Alzheimer's disease pathology and microglial activation are associated with altered white matter microstructure in asymptomatic adults at risk for Alzheimer's disease. *J Alzheimers Dis* **50**, 873-886.

[39] Antonell A, Mansilla A, Rami L, Lladó A, Iranzo A, Olives J, Balasa M, Sánchez-Valle R, Molinuevo JL (2014) Cerebrospinal fluid level of YKL-40 protein in preclinical and prodromal Alzheimer's disease. *J Alzheimers Dis* **42**, 901-908.

[40] Verdile G, Gnjec A, Miklossy J, Fonte J, Veurink G, Bates K, Kakulas B, Mehta PD, Milward EA, Tan N, Lareu R, Lim D, Dharmarajan A, Martins RN (2004) Protein markers for Alzheimer disease in the frontal cortex and cerebellum. *Neurology* **63**, 1385-1392.

[41] Scahill RI, Ridgway GR, Bartlett JW, Barnes J, Ryan NS, Mead S, Beck J, Clarkson MJ, Crutch SJ, Schott JM, Ourselin S, Warren JD, Hardy J, Rossor MN, Fox NC (2013) Genetic influences on atrophy patterns in familial alzheimer's disease: A comparison of APP and PSEN1 mutations. *J Alzheimers Dis* **35**, 199-212.

[42] Knight WD, Kim LG, Douiri A, Frost C, Rossor MN, Fox NC (2011) Acceleration of cortical thinning in familial Alzheimer's disease. *Neurobiol Aging* **32**, 1765-1773.

[43] Balasa M, Vidal-Piñeiro D, Lladó A, Antonell A, Bosch B, Castellanos F, Bargalló N, Bartres-Faz D, Molinuevo JL, Sánchez-Valle R (2012) PSEN1 mutation carriers present lower cerebrospinal fluid amyoid-β42 levels than sporadic early-onset Alzheimer's disease patients but no differences in neuronal injury biomarkers. *J Alzheimers Dis* **30**, 605-616.

[44] Gutiérrez-Galve L, Lehmann M, Hobbs NZ, Clarkson MJ, Ridgway GR, Crutch S, Ourselin S, Schott JM, Fox NC, Barnes J (2009) Patterns of cortical thickness according to APOE genotype in Alzheimer's disease. *Dement Geriatr Cogn Disord* **28**, 476-485.

[45] Donix M, Burggren AC, Scharf M, Marschner K, Suthana NA, Siddarth P, Krupa AK, Jones M, Martin-Harris L, Ercoli LM, Miller KJ, Werner A, von Kummer R, Sauer C, Small GW, Holthoff VA, Bookheimer SY (2013) APOE associated hemispheric asymmetry of entorhinal cortical thickness in aging and Alzheimer's disease. *Psychiatry Res* **214**, 212-220.

[46] Yoo H, Lee H, Shin S, Park S-W, Choi J, Jung H, Cha J, Lee J-M, Lee J-Y (2014) Butyrylcholinesterase K and apolipoprotein Ɛ4 affect cortical thickness and neuropsychiatric symptoms in Alzheimer's disease. *Curr Alzheimer Res* **11**, 137-144.

[47] Foley JM, Salat DH, Stricker NH, McGlinchey RE, Milberg WP, Grande LJ, Leritz EC (2016) Glucose dysregulation interacts with APOE- Ɛ4 to potentiate temporoparietal cortical thinning. *Am J Alzheimers Dis Other Demen* **31**, 76-86.

[48] Mukherjee S, Kim S, Gibbons LE, Nho K, Risacher SL, Glymour MM, Habeck C, Lee GJ, Mormino E, Ertekin-Taner N, Montine TJ, Decarli C, Saykin AJ, Crane PK (2012) Genetic architecture of resilience of executive functioning. *Brain Imaging Behav* **6**, 621-633.

[49] Laczó J, Andel R, Vyhnalek M, Matoska V, Kaplan V, Nedelska Z, Lerch O, Gazova I, Moffat SD, Hort J (2015) The effect of TOMM40 on spatial navigation in amnestic mild cognitive impairment. *Neurobiol Aging* **36**, 2024-2033.

[50] Yang X, Liu P, Sun J, Wang G, Zeng F, Yuan K, Liu J, Dong M, von Deneen KM, Qin W, Tian J (2012) Impact of brain-derived neurotrophic factor Val66Met polymorphism on cortical thickness and voxel-based morphometry in healthy Chinese young adults. *PLoS One* **7**, e37777.

[51] Zugman A, Pedrini M, Gadelha A, Kempton MJ, Noto CS, Mansur RB, Santoro ML, Gama CS, Bressan RA, McGuire P, Jackowski AP, Brietzke E (2015) Serum brain-derived neurotrophic factor and cortical thickness are differently related in patients with schizophrenia and controls. *Psychiatry Res* **234**, 84-89.

[52] Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, Cruchaga C, Sassi C, Kauwe JSK, Younkin S, Hazrati L, Collinge J, Pocock J, Lashley T, Williams J, Lambert J-C, Amouyel P, Goate A, Rademakers R, Morgan K, Powell J, St George-Hyslop P, Singleton A, Hardy J, Alzheimer Genetic Analysis Group (2013) TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117-127.

[53] Zhong L, Chen XF, Zhang ZL, Wang Z, Shi XZ, Xu K, Zhang YW, Xu H, Bu G (2015) DAP12 stabilizes the C-terminal fragment of the triggering receptor expressed on myeloid cells-2 (TREM2) and protects against LPS-induced pro-inflammatory response. *J Biol Chem* **290**, 15866-15877.

[54] Průcha M, Zazula R, Müller M, Hyánek T, Dostál M, Sedláčková L (2011) TREM-1 expression on monocytes is not a parameter specific for infectious etiology of systemic inflammatory response syndrome. *Prague Med Rep* **112**, 205-215.

[55] Rohn TT (2013) The triggering receptor expressed on myeloid cells 2: "TREM-ming" the inflammatory component associated with Alzheimer's disease. *Oxid Med Cell Longev* **2013**, 860959.

[56] Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, Bjornsson S, Huttenlocher J, Levey AI, Lah JJ, Rujescu D, Hampel H, Giegling I, Andreassen OA, Engedal K, Ulstein I, Djurovic S, Ibrahim-Verbaas C, Hofman A, Ikram MA, van Duijn CM, Thorsteinsdottir U, Kong A, Stefansson K (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368**, 107-116.

[57] Luis EO, Ortega-Cubero S, Lamet I, Razquin C, Cruchaga C, Benitez BA, Lorenzo E, Irigoyen J, Pastor MA, Pastor P (2014) Frontobasal gray matter loss is associated with the TREM2 p. *R47H variant Neurobiol Aging* **35**, 2681-2690.

[58] Lue LF, Schmitz CT, Serrano G, Sue LI, Beach TG, Walker DG (2015) TREM2 protein expression changes correlate with Alzheimer's disease neurodegenerative pathologies in post-mortem temporal cortices. *Brain Pathol* **25**, 469-480.

[59] Tanaka S, Kawamata J, Shimohama S, Akaki H, Akiguchi I, Kimura J, Ueda K (1998) Inferior temporal lobe atrophy and APOE genotypes in Alzheimer's disease. X-ray computed tomography, magnetic resonance imaging and Xe-133 SPECT studies. *Dement Geriatr Cogn Disord* **9**, 90-98.

[60] Tai LM, Thomas R, Marottoli FM, Koster KP, Kanekiyo T, Morris AWJ, Bu G (2016) The role of APOE in cerebrovascular dysfunction. *Acta Neuropathol* **131**, 709-723.

[61] Moreno-Ortega AJ, Ruiz-Nuño A, García AG, Cano-Abad MF (2010) Mitochondria sense with different kinetics the calcium entering into HeLa cells through calcium channels CALHM1 and mutated P86L-CALHM1. *Biochem Biophys Res Commun* **391**, 722-726.

[62] Vingtdeux V, Chandakkar P, Zhao H, Blanc L, Ruiz S, Marambaud P (2015) CALHM1 ion channel elicits amyloid- clearance by insulin-degrading enzyme in cell lines and in vivo in the mouse brain. *J Cell Sci* **128**, 2330-2338.

[63] Yao J, Irwin RW, Zhao L, Nilsen J, Hamilton RT, Brinton RD (2009) Mitochondrial bioenergetic deficit precedes Alzheimer's pathology in female mouse model of Alzheimer's disease. *Proc Natl Acad Sci U S A* **106**, 14670-14675.

[64] Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, Foroud T, Pankratz N, Moore JH, Sloan CD, Huentelman MJ, Craig DW, DeChairo BM, Potkin SG, Jack CR, Weiner MW, Saykin AJ (2010) Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* **53**, 1051-1063.

[65] Carmona MA, Murai KK, Wang L, Roberts AJ, Pasquale EB (2009) Glial ephrin-A3 regulates hippocampal dendritic spine morphology and glutamate transport. *Proc Natl Acad Sci U S A* **106**, 12524-12529.

[66] Braskie MN, Ringman JM, Thompson PM (2011) Neuroimaging measures as endophenotypes in Alzheimer's disease. *Int J Alzheimers Dis* **2011**, 1-15.

[67] Pawlowski JW, Kellicker N, Bobst CE, Kaltashov I (2016) Assessing the iron delivery efficacy of transferrin in clinical samples by native electrospray ionization mass spectrometry. *Analyst* **141**, 853-861.

[68] Printy BP, Verma N, Cowperthwaite MC, Markey MK (2014) Effects of genetic variation on the dynamics of neurodegeneration in Alzheimer's disease. *Conf Proc IEEE Eng Med Biol Soc* **2014**, 2464-2467.

[69] Obirikorang C, Issahaku RG, Osakunor DNM, Osei-Yeboah J (2016) Anaemia and iron homeostasis in a cohort of HIV-infected patients: A cross-sectional study in Ghana. *AIDS Res Treat* **2016**, 1623094.

[70] Silver JM, McAllister TW, Yudofsky SC (2005) *Textbook of traumatic brain injury*. American Psychiatric Pub.

[71] Ceusters W, Smith B (2006) A realism-based approach to the evolution of biomedical ontologies. *AMIA Annu Symp Proc* **2006**, 121-125.

[72] Amdouni E, Morvan Y, Gibaud B (2015) Towards an imaging biomarker ontology based on the open biological and biomedical ontologies foundry. In *Proceedings Semantic Web Applications and Tools for Life Sciences*, pp. 2-3.

[73] Zhang L, Chang RC, Chu L, Mak HK (2012) Current neuroimaging techniques in Alzheimer's disease and applications in animal models. *Am J Nucl Med Mol Imaging* **2**, 386-404.

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

[75] Younesi E, Malhotra A, Gündel M, Scordis P, Kodamullil AT, Page M, Müller B, Springstubbe S, Wüllner U, Scheller D, Hofmann-Apitius M (2015) PDON: Parkinson's disease ontology for representation and modeling of the Parkinson's disease knowledge domain. *Theor Biol Med Model* **12**, 20.

[76] Malhotra A, Gündel M, Rajput AM, Mevissen HT, Saiz A, Pastor X, Lozano-Rubi R, Martinez-Lapsicina EH, Zubizarreta I, Mueller B, Kotelnikova E, Toldo L, Hofmann-Apitius M, Villoslada P (2015) Knowledge retrieval from pubmed abstracts and electronic medical records with the multiple sclerosis ontology. *PLoS One* **10**, 1-12.

[77] Dreses-Werringloer U, Lambert JC, Vingtdeux V, Zhao H, Vais H, Siebert A, Jain A, Koppel J, Rovelet-Lecrux A, Hannequin D, Pasquier F, Galimberti D, Scarpini E, Mann D, Lendon C, Campion D, Amouyel P, Davies P, Foskett JK, Campagne F, Marambaud P (2008) A polymorphism in CALHM1 influences Ca2+homeostasis, Aβ levels, and Alzheimer's disease risk. *Cell* **133**, 1149-1161.

[78] Schupf N, Lee A, Park N, Dang L-H, Pang D, Yale A, Oh DK-T, Krinsky-McHale SJ, Jenkins EC, Luchsinger JA,

Zigman WB, Silverman W, Tycko B, Kisselev S, Clark L, Lee JH (2015) Candidate genes for Alzheimer's disease are associated with individual differences in plasma levels of beta amyloid peptides in adults with Down syndrome. *Neurobiol Aging* **36**, 2907.e1-2907.e10.

[79] Deming Y, Black K, Carrell D, Cai Y, Del-Aguila JL, Fernandez MV, Budde J, Ma S, Saef B, Howells B, Bertelsen S, Huang K, Sutphen CL, Tarawneh R, Fagan AM, Holtzman DM, Morris JC, Goate AM, Dougherty JD, Cruchaga C (2016) Chitinase-3-like 1 protein (CHI3L1) locus influences cerebrospinal fluid levels of YKL-40. *BMC Neurol* **16**, 217.

[80] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH (2012) Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 280-292.

[81] Nelson PT, Braak H, Markesbery WR (2009) Neuropathology and cognitive impairment in Alzheimer disease: A complex but coherent relationship. *J Neuropathol Exp Neurol* **68**, 1-14.

SUMMARY

There is ample clinical evidence to support the fact that the pathophysiological onset of AD commences much earlier than the clinical manifestation of the disease. However, current diagnostic and therapeutic approaches are only symptomatic; therefore, they are not capable yet to slow cognitive decline or the pathological progression. With the recent advent of neuroimaging techniques and their capacity to provide non-invasive, high-resolution information about the physiological and anatomical aspects of the brain, this measurable outcome can play a significant role in clinical trials for early diagnosis and treatment. Although current disease-modifying treatments are based on cognitive and behavioral assessments, such as ADAS-Cog or the caretaker's observation of the patient, these measurements are not yet robust enough to determine the reason for cognitive decline or alter behavior, which can differ from patient-to-patient. This is because such trials require large patient cohorts and continuous assessments of patients for a minimum of five years.

Unlike other biomarkers, neuroimaging biomarkers can be measured much more frequently and are highly non-invasive, enabling them to track the progression of the disease across varying stages. These imaging biomarkers are invaluable in linking and validating existing molecular hypotheses with clinical outcome measures of the disease, making them widely prevalent in the field of neurodegeneration.

Nonetheless, with the growing popularity of imaging technologies in scientific research, the amount of knowledge derived and data produced from multiple experiments are primarily on the rise. This has eventually given rise to independently-developed platforms for collecting, describing and storing data, resulting in the lack

of a standardized and reliable framework for reusing the generated data. Furthermore, the data derived from raw image scans are not human readable.

Biomedical ontologies have often proven to be valuable in extracting domain relevant information from brain image scans, and other software generated imaging readouts. However, the amount of knowledge and information published in the literature based on imaging analysis often goes unnoticed.

Moreover, neuroimaging readouts are predominantly used in disease-specific investigations for tracking or observing disease progression of individuals/groups of patients. For this purpose, the NIFT terminology was constructed to represent, integrate and harmonize "neuroimaging pointers/readouts" in the context of NDD with high specificity and granularity. This terminology provides a broad scope of usage in NDD research for efficient data interoperability and exchange amongst clinicians and scientific researchers making it a gold standard reference for apprehending clinical readouts from brain scans. The next aspect of the publication highlights on how relevant knowledge extracted using the NIFT terminology can be successfully integrated into a knowledge model for deriving a deeper understanding of mechanistic processes across all scales involved in the disease. The final aspect of the manuscript focuses on the integration and formalization of "big data" such as ADNI and PPMI, which are largest cohort, based longitudinal studies for efficient data sharing of patient-centric data. With access to real-time longitudinal patient data, it is now possible to garner a realistic understanding of pathophysiological processes involved in NDD and find effective treatment therapies, which can be custom-made for patients.

CHAPTER 4

Conclusion and Outlook

Alzheimer's disease (AD) is estimated to quadruple in the coming years, posing a severe challenge to the healthcare community for early prevention and treatment. Much of the research in the field of neurodegeneration has traditionally been conducted on animal-based experiments with the intention of gaining enhanced insights into the pathophysiological mechanisms of the disease. However, it is increasingly becoming evident that earlier research efforts have not been successful in translating scientific findings into successful therapies and treatments for patients suffering from NDDs. The translational failure of these approaches implicates that the sporadic form of AD is multifactorial, requiring the intervention of experts coming from multiple research domains. The advancement of technology in recent years has empowered the research world, particularly in the field of NDD, to generate, collect and store massive amounts of data in disparate data repositories. However, the problem of data redundancy, as well as a lack of standardized means for data representation, has made it nearly impossible to exploit the full potential of the data resources effectively.

4.1.1 Need for shared semantics

The first section of this dissertation is dedicated to advocating the need for a more comprehensive approach when dealing with large-scale heterogeneous datasets to

elucidate a better understanding of disease pathology. The heterogeneity amongst various data resources as well as a lack of transparency has been hindrances for the research community for many decades. The evolution of semantic web technology has fostered the possibility of integrating heterogeneous data through a consistent framework but has also enabled mining "useful" information from the embedded resources. The lack of semantic web approaches dedicated to neurodegeneration research resulted in the high-resolution, semantic framework known as NeuroRDF, a platform for asking complex, scientific questions on neurological diseases.

4.1.2 Effect of dominant ontologies and terminologies for enriching knowledge models

The second section of this dissertation is directed towards the vital usability of ontologies and terminologies to aggregate scattered knowledge from literature. The systems biology community is often challenged by the massive influx of research publications in PubMed. The text-mining community has primarily grown over the years by providing systematic and automated support for extracting domain-specific knowledge from the literature. Although such terminologies and ontologies are useful for extracting valuable insights from the literature, they still do not contribute directly towards understanding the pathomechanisms of a particular disease.

During this thesis, a pathway terminology system (PTS) terminology was developed to extract disease-specific, pertinent pathway information from the literature, particularly for NDD. With the help of this terminology, it is now possible to integrate

pathway relevant knowledge from the literature into computable models, thereby enriching and adding value to existing knowledge resources.

4.1.3 Clinical data – a potential goldmine for deriving novel hypothesis

The third section of this dissertation is devoted toward exploring the unchartered territories of clinical data in the field of NDD. Data privacy and ethical boundaries have often hindered the research community from getting access to real-time patient data. With the growing recognition of the open source framework community, the need for data sharing is becoming of utmost importance, especially in the field of NDD. Despite the encouragement and facilitation of ongoing research efforts as well as funding agencies, the medical domain remains inaccessible due to ethical, political as well as technical barriers that prevent them from outreach to the research community.

Although the fruits of data sharing were widely acknowledged by the scientific community, only after the boom of global initiatives like ADNI, PPMI, and BIRN, was the idea of comprehensive data sharing cemented. The free accessibility of "big data" has opened up possibilities for various scientific organizations to come together to solve specific biological problems, such as the big data DREAM challenge. The goal of the Alzheimer's Disease Big Data DREAM was to apply an open science approach to rapidly identify accurate predictive AD biomarkers that could be used by the scientific, industrial and regulatory communities to improve AD diagnosis and treatment. It was also the first in a series of AD Data Challenges to leverage genetics and brain imaging in combination with cognitive assessments, biomarkers and demographic information from cohorts ranging from cognitively normal to mild cognitively impaired to individuals with AD.

Although the challenge overwhelmed the community with the amount of data gathered, it also gave insight into two important aspects:

(i) The quality of data present which was not harmonized or curated across different scales and measurements

(ii) The need for bridging the gap between molecular and genetic factors with clinical and radiological features

The lack of order in the representation of clinical data led to the construction of a framework called NIFT to assist in standardizing the diagnostic as well as measured indices observed using neuroimaging techniques. NIFT not only supports organization and exchange of imaging information among neurologists and clinical researchers but is also a useful tool for annotation of brain scan metadata.

Importantly, the usage of NIFT is not only restricted to data retrieval and extraction, but it also acts as a potential tool for the neuroscience community to bridge the gap between scales ranging from the level of molecular mechanisms to imaging readouts. This can only be achieved through a contextual and mechanistic understanding between the latter two scales, rather than association-based studies, which are statistically significant, but lack in providing this more in-depth understanding. For this purpose, a computational model was constructed around the molecular features and clinical outcomes. This type of integration across scales is often regarded as the "holy grail" of integrative modeling and mining.

4.2 Outlook

The goal of this thesis was, (i) to demonstrate the possibility of harmonizing context-specific knowledge resources and data derivatives into a semantic framework to facilitate the identification of potential biomarkers in NDD, (ii) to organize and structure pathway relevant information from the literature and to use them for various data mining applications, and (iii) to initiate an effort in exploring and organizing clinically relevant knowledge and determine their potential role in deciphering the disease pathophysiology of NDD. The efforts undertaken during this dissertation have only begun to reveal the possibilities for dementia treatment and intervention. With the growing appreciation for sharing big data from the medical community, the future of neuroscience research, particularly for AD, appears to be promising. The availability of patient-specific longitudinal data from ADNI is revolutionary as it opens new doors for building stage-specific longitudinal disease models using real-time patient data.

One futuristic approach with real-time patient data could be directed towards generating virtual patient cohorts for improvising personalized care and treatment for patients. Computational models are often considered a unique platform for providing customized solutions, and when they can be coupled with mathematical models, they can often lead to simulations of real-time patients into virtual patients with varying inter and intrapatient parameters. Such combinatorial approaches could be used for validating in-silico approaches for testing different protocols without restrictive legal and ethical conflicts.

REFERENCES

[1] Herculano-Houzel S (2009) The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.* **3**, 1–11.

[2] Betzel RF, Bassett DS (2017) Multi-scale brain networks. *Neuroimage* **160**, 73–83.

[3] Bassett DS The Brain : A Multiscale System.

[4] Crosson B, Ford A, McGregor KM, Meinzer M, Cheshkov S, Li X, Walker-Batson D, Briggs RW (2010) Functional imaging and related techniques: An introduction for rehabilitation researchers. *J. Rehabil. Res. Dev.* **47**, vii.

[5] Wang J, Zuo X, Dai Z, Xia M, Zhao Z, Zhao X, Jia J, Han Y, He Y (2013) Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. *Biol. Psychiatry* **73**, 472–481.

[6] Atasoy S, Donnelly I, Pearson J (2016) Human brain networks function in connectome-specific harmonic waves. *Nat. Commun.* **7**, 1–10.

[7] Sporns O (2013) Structure and function of complex brain networks. *Dialogues Clin. Neurosci.* **15**, 247–262.

[8] Sethi SS, Zerbi V, Wenderoth N, Fornito A, Fulcher BD (2017) Structural connectome topology relates to regional BOLD signal dynamics in the mouse brain. *Chaos* **27**,.

[9] Burden AP, Prescott JW (2014) Connectome : Changes in Cortical Network Topology with Increased. **273**, 175–184.

[10] Contreras JA, Goni J, Risacher SL, Sporns O, Saykin AJ (2015) The Structural and Functional Connectome and Prediction of Risk for Cognitive Impairment in Older Adults. *Curr. Behav. Neurosci. Reports* **2**, 234–245.

[11] Julia M (2017) 1,2,3 ,.

[12] Liang H, Wang H (2017) Structure-Function Network Mapping and Its Assessment via Persistent Homology. *PLoS Comput. Biol.* **13**, 1–19.

[13] Shimono M, Beggs JM (2015) Functional clusters, hubs, and communities in the cortical microconnectome. *Cereb. Cortex* **25**, 3743–3757.

[14] Bressler SL, Menon V (2010) Large-scale brain networks in cognition: emerging

methods and principles. *Trends Cogn. Sci.* **14**, 277–290.

[15] Philippi N, Rousseau F, Noblet V, Botzung A, Després O, Cretin B, Kremer S, Blanc F, Manning L (2015) Different Temporal Patterns of Specific and General Autobiographical Memories across the Lifespan in Alzheimer's Disease. *Behav. Neurol.* **2015**,.

[16] Echtermeyer C, Han CE, Rotarska-Jagiela A, Mohr H, Uhlhaas PJ, Kaiser M (2011) Integrating Temporal and Spatial Scales: Human Structural Network Motifs Across Age and Region of Interest Size. *Front. Neuroinform.* **5**, 1–14.

[17] Singh SS, Khundrakpam B, Reid AT, Lewis JD, Evans AC, Ishrat R, Sharma BI, Singh RKB (2016) Scaling in topological properties of brain networks. *Sci. Rep.* **6**, 1–19.

[18] Przedborski S, Vila M, Jackson-lewis V (2003) Neurodegeneration : What is it and where are we ? *J. Clin. Invest.* **111**, 3–10.

[19] Wimo A, Guerchet M, Ali GC, Wu YT, Prina AM, Winblad B, Jönsson L, Liu Z, Prince M (2017) The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's Dement.* **13**, 1–7.

[20] Karlawish J (2014) How are we going to live with Alzheimer's disease? *Health Aff.* **33**, 541–546.

[21] Prince M, Wimo A, Guerchet M, Ali GC, Wu Y-T, Prina M (2015) World Alzheimer Report 2015 - The Global Impact of Dementia.

[22] Fox M, Knapp LA, Andrews PW, Fincher CL (2013) Hygiene and the world distribution of Alzheimer's disease: Epidemiological evidence for a relationship between microbial environment and age-adjusted disease burden. *Evol. Med. public Heal.* **2013**, 173–86.

[23] Khan T (2016) *Biomarkers in Alzheimer's Disease*.

[24] Dharmarajan TS, Gunturu SG (2009) Alzheimer's disease: A healthcare burden of epidemic proportion. *Am. Heal. Drug Benefits* **2**, 39–47.

[25] Au R, Piers RJ, Lancashire L (2015) Back to the future: Alzheimer's disease

heterogeneity revisited. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **1**, 368–370.

[26] Willén K, Sroka A, Takahashi RH, Gouras GK (2017) Heterogeneous Association of Alzheimer's Disease-Linked Amyloid-β and Amyloid-β Protein Precursor with Synapses. *J. Alzheimer's Dis.* **60**, 511–524.

[27] The Lancet (2016) Alzheimer's disease: expedition into the unknown. *Lancet* **388**, 2713.

[28] Cummings J, Lee G, Mortsdorf T, Ritter A, Zhong K (2017) Alzheimer's disease drug development pipeline: 2017. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **3**, 367–384.

[29] Kamboh MI, Demirci FY, Wang X, Minster RL, Carrasquillo MM, Pankratz VS, Younkin SG, Saykin AJ, Jun G, Baldwin C, Logue MW, Buros J, Farrer L, Pericak-Vance MA, Haines JL, Sweet RA, Ganguli M, Feingold E, Dekosky ST, Lopez OL, Barmada MM (2012) Genome-wide association study of Alzheimer's disease. *Transl. Psychiatry* **2**, 1–7.

[30] Galvin JE (2017) Prevention of Alzheimer's Disease: Lessons Learned and Applied. *J. Am. Geriatr. Soc.* **65**, 2128–2133.

[31] Robinson M, Lee BY, Hane FT (2017) Recent Progress in Alzheimer's Disease Research, Part 2: Genetics and Epidemiology. *J. Alzheimer's Dis.* **57**, 317–330.

[32] Scheff SW, Ginsberg SD (2017) Alzheimer ' s Disease. 54–69.

[33] Higham JP (2017) Tangled tau: Active pathology or footprint of disease? *Alzheimer's Dement. Transl. Res. Clin. Interv.* **3**, 658–659.

[34] Khan SS, Bloom GS (2016) Tau: The Center of a Signaling Nexus in Alzheimer's Disease. *Front. Neurosci.* **10**, 1–5.

[35] Bronzuoli MR, Iacomino A, Steardo L, Scuderi C (2016) Targeting neuroinflammation in Alzheimer's disease. *J. Inflamm. Res.* **9**, 199–208.

[36] Liu AKL, Goldfinger MH, Questari HE, Pearce RKB, Gentleman SM (2016) ARTAG in the basal forebrain: widening the constellation of astrocytic tau pathology. *Acta*

Neuropathol. Commun. **4**, 59.

[37] Crary JF (2016) Primary age-related tauopathy and the amyloid cascade hypothesis: the exception that proves the rule? *J. Neurol. neuromedicine* **1**, 53–57.

[38] Kim BM, You MH, Chen CH, Lee S, Hong Y, Hong Y, Kimchi A, Zhou XZ, Lee TH (2014) Death-associated protein kinase 1 has a critical role in aberrant tau protein regulation and function. *Cell Death Dis.* **5**, e1237-12.

[39] Fick DM, Kolanowski A, Beattie E, McCrow J (2009) Delirium in early-stage alzheimer's disease: enhancing cognitive reserve as a possible preventive measure. *J. Gerontol. Nurs.* **35**, 30–8.

[40] Harris D (2007) Forget me not: Palliative care for people with dementia. *Postgrad. Med. J.* **83**, 362–366.

[41] Tolea MI, Morris JC, Galvin JE (2016) Trajectory of mobility decline by type of dementia. *Alzheimer Dis. Assoc. Disord.* **30**, 60–66.

[42] Iqbal K, Grundke-Iqbal I (2010) Alzheimer's disease, a multifactorial disorder seeking multitherapies. *Alzheimer's Dement.* **6**, 420–424.

[43] Carrettiero DC, Santiago FE, Motzko-Soares ACP, Almeida MC (2015) Temperature and toxic Tau in Alzheimer's disease: new insights. *Temp. (Austin, Tex.)* **2**, 491–8.

[44] Rao AT, Degnan AJ, Levy LM (2014) Genetics of Alzheimer disease. *Am. J. Neuroradiol.* **35**, 457–458.

[45] Lanoiselée HM, Nicolas G, Wallon D, Rovelet-Lecrux A, Lacour M, Rousseau S, Richard AC, Pasquier F, Rollin-Sillaire A, Martinaud O, Quillard-Muraine M, de la Sayette V, Boutoleau-Bretonniere C, Etcharry-Bouyx F, Chauviré V, Sarazin M, le Ber I, Epelbaum S, Jonveaux T, Rouaud O, Ceccaldi M, Félician O, Godefroy O, Formaglio M, Croisile B, Auriacombe S, Chamard L, Vincent JL, Sauvée M, Marelli-Tosi C, Gabelle A, Ozsancak C, Pariente J, Paquet C, Hannequin D, Campion D (2017) APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. *PLoS Med.* **14**, 1–16.

[46] Cacace R, Sleegers K, Van Broeckhoven C (2016) Molecular genetics of early-onset

Alzheimer's disease revisited. *Alzheimer's Dement.* **12**, 733–748.

[47] Young AL, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, Schott JM, Alexander DC (2014) A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain* **137**, 2564–2577.

[48] de la Monte SM, Wands JR (2008) Alzheimer's Disease Is Type 3 Diabetes — Evidence Reviewed. *J. Diabetes Sci. Technol.* **2**, 1101–1113.

[49] Ryan NS, Rossor MN (2010) Correlating familial Alzheimer's disease gene mutations with clinical phenotype. *Biomark. Med.* **4**, 99–112.

[50] Folch J, Ettcheto M, Petrov D, Abad S, Pedrós I, Marin M, Olloquequi J, Camins A (2017) Review of the advances in treatment for Alzheimer disease: strategies for combating β-amyloid protein. *Neurol. (English Ed.*

[51] Fjell AM, McEvoy L, Holland D, Dale AM, Walhovd KB (2014) What is normal in normal aging? Effects of Aging, Amyloid and Alzheimer's Disease on the Cerebral Cortex and the Hippocampus. *Prog. Neurobiol.* **117**, 20–40.

[52] Selkoe DJ, Hardy J (2016) The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8**, 595–608.

[53] Maarouf CL, Daugs ID, Kokjohn TA, Walker DG, Hunter JM, Kruchowsky JC, Woltjer R, Kaye J, Castaño EM, Sabbagh MN, Beach TG, Roher AE (2011) Alzheimer's disease and non-demented high pathology control nonagenarians: Comparing and contrasting the biochemistry of cognitively successful aging. *PLoS One* **6**, 1–17.

[54] Briley D, Ghirardi V, Woltjer R, Renck A, Zolochevska O, Taglialatela G, Micci MA (2016) Preserved neurogenesis in non-demented individuals with AD neuropathology. *Sci. Rep.* **6**, 1–10.

[55] Geerts H, Dacks PA, Devanarayan V, Haas M, Khachaturian ZS, Gordon MF, Maudsley S, Romero K, Stephenson D (2016) Big data to smart data in Alzheimer's disease: The brain health modeling initiative to foster actionable knowledge. *Alzheimer's Dement.* **12**, 1014–1021.

[56] Kaiser NC, Ph D, Liang L, Wilkins SS, Sultzer DL, Mendez MF (2014) Among Patients With Alzheimer ' s Disease. 73–80.

[57] Folch, J., Petrov, D., Ettcheto, M., Abad, S., Sanchez-Lopez, E., Garcia, M.L., Olloquequi, J., Beas-Zarate, C., Auladell, C., Antoni C (2015) Current Research Therapeutic Strategies for Alzheimer's Disease Treatment. *Neural Plast.* **2016**, 15.

[58] Iqbal K, Chohan MO, Grundke-Iqbal I (2008) Stratification of Patients is the Way to Go to Develop Neuroprotective/Disease-Modifying Drugs for Alzheimer's Disease. *J Alzheimers Dis* **15**, 339–345.

[59] Sharma N, Singh AN (2016) Exploring biomarkers for Alzheimer's disease. *J. Clin. Diagnostic Res.* **10**, KE01-KE06.

[60] Merlo Pich E, Jeromin A, Frisoni GB, Hill D, Lockhart A, Schmidt ME, Turner MR, Mondello S, Potter WZ (2014) Imaging as a biomarker in drug discovery for Alzheimer's disease: Is MRI a suitable technology? *Alzheimer's Res. Ther.* **6**, 1–7.

[61] Ritter A, Cummings J (2015) Fluid biomarkers in clinical trials of Alzheimer's disease therapeutics. *Front. Neurol.* **6**,.

[62] Niemantsverdriet E, Valckx S, Bjerke M, Engelborghs S (2017) Alzheimer's disease CSF biomarkers: clinical indications and rational use. *Acta Neurol. Belg.* **117**, 591–602.

[63] Caroli A, Galluzzi S, Ferrari C, Prestia A, Van Der Flier WM, Ossenkoppele R, Van Berckel BNM, Barkhof F, Teunissen CE, Wall A, Carter SF, Scholl M, Choo IH, Grimmer T, Redolfi A, Nordberg A, Scheltens P, Drzezga A, Frisoni GB (2015) Alzheimer's disease core biomarkers and prediction of dementia in MCI: The effect of age at onset. *Alzheimer's Dement.* **11**, P140–P142.

[64] Huynh RA, Mohan C (2017) Alzheimer's disease: Biomarkers in the genome, blood, and cerebrospinal fluid. *Front. Neurol.* **8**,.

[65] Anoop A, Singh PK, Jacob RS, Maji SK (2010) CSF Biomarkers for Alzheimer's Disease Diagnosis. *Int. J. Alzheimers. Dis.* **2010**, 1–12.

[66] Dekker AD, Fortea J, Blesa R, De Deyn PP (2017) Cerebrospinal fluid biomarkers for

Alzheimer's disease in Down syndrome. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **8**, 1–10.

[67] Armon C, Evans RW (2005) Addendum to assessment: Prevention of post-lumbar puncture headaches: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* **65**, 510–512.

[68] E. P, a. N, T. D-S, H. S (2009) Safety of lumbar puncture procedures in patients with Alzheimer's disease. *Curr. Alzheimer Res.* **6**, 290–292.

[69] Thambisetty M, Lovestone S (2010) Blood-based biomarkers of Alzheimer's disease: challenging but feasible. *Biomark. Med.* **4**, 65–79.

[70] Kaj Blennowa,*, Bruno Duboisb, Anne M. Faganc, Piotr Lewczukd, Mony J. de Leone, f, And Hampel H (2015) Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's disease Kaj. *HHS Public Access* **11**, 58–69.

[71] Skillbäck T, Farahmand BY, Rosén C, Mattsson N, Nägga K, Kilander L, Religa D, Wimo A, Winblad B, Schott JM, Blennow K, Eriksdotter M, Zetterberg H (2015) Cerebrospinal fluid tau and amyloid-β<inf>1-42</inf> in patients with dementia. *Brain* **138**, 2716–2731.

[72] Mo J-A, Lim J-H, Sul A-R, Lee M, Youn YC, Kim H-J (2015) Cerebrospinal Fluid β-Amyloid1–42 Levels in the Differential Diagnosis of Alzheimer's Disease—Systematic Review and Meta-Analysis. *PLoS One* **10**, e0116802.

[73] Chow VW, Mattson MP, Wong PC, Gleichmann M (2011) Neuronal Activity and the Expression of Clathrin Assembly Protein AP180. *Biochem Biophys Res Commun* **12**, 1–12.

[74] Haass C, Kaether C, Thinakaran G, Sisodia S (2012) Trafficking and proteolytic processing of APP. *Cold Spring Harb. Perspect. Med.* **2**, 1–25.

[75] Forlenza O V., Radanovic M, Talib LL, Aprahamian I, Diniz BS, Zetterberg H, Gattaz WF (2015) Cerebrospinal fluid biomarkers in Alzheimer's disease: Diagnostic accuracy and prediction of dementia. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **1**, 455–463.

[76] Fagan AM, Holtzman DM (2010) Cerebrospinal fluid biomarkers of Alzheimer's disease. *Biomark. Med.* **4**, 51–63.

[77] Randall C, Mosconi L, de Leon M, Glodzik L (2013) Cerebrospinal fluid biomarkers of Alzheimer's disease in healthy elderly. *Front. Biosci. (Landmark Ed.* **18**, 1150–73.

[78] Gilmore JH (2008) NIH Public Access. *North* **29**, 1883–1889.

[79] Lue L-F, Guerra A, Walker DG (2017) Amyloid Beta and Tau as Alzheimer's Disease Blood Biomarkers: Promise From New Technologies. *Neurol. Ther.* **6**, 25–36.

[80] Ruiz A, Pesini P, Espinosa A, Pérez-Grijalba V, Valero S, Sotolongo-Grau O, Alegret M, Monleón I, Lafuente A, Buendía M, Ibarria M, Ruiz S, Hernández I, San José I, Tárraga L, Boada M, Sarasa M (2013) Blood amyloid beta levels in healthy, mild cognitive impairment and Alzheimer's disease individuals: Replication of diastolic blood pressure correlations and analysis of critical covariates. *PLoS One* **8**, 1–9.

[81] Hitoshi Sohma YK (2016) *Plasma Biomarkers in Alzheimer's Disease*.

[82] Kumar P, Dezso Z, MacKenzie C, Oestreicher J, Agoulnik S, Byrne M, Bernier F, Yanagimachi M, Aoshima K, Oda Y (2013) Circulating miRNA Biomarkers for Alzheimer's Disease. *PLoS One* **8**,.

[83] Muniategui A, Nogales-Cadenas R, Vázquez M, Aranguren XL, Agirre X, Luttun A, Prosper F, Pascual-Montano A, Rubio A (2012) Quantification of miRNA-mRNA interactions. *PLoS One* **7**,.

[84] Yeri A, Courtright A, Reiman R, Carlson E, Beecroft T, Janss A, Siniard A, Richholt R, Balak C, Rozowsky J, Kitchen R, Hutchins E, Winarta J, McCoy R, Anastasi M, Kim S, Huentelman M, Van Keuren-Jensen K (2017) Total extracellular small RNA profiles from plasma, saliva, and urine of healthy subjects. *Sci. Rep.* **7**, 1–13.

[85] Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, Galas DJ, Wang K (2016) HHS Public Access. **56**, 1733–1741.

[86] Tarawneh R, Holtzman DM (2012) The clinical problem of symptomatic Alzheimer disease and mild cognitive impairment. *Cold Spring Harb. Perspect. Med.* **2**, 1–16.

[87] Blair RJR (2015) Clinical research. *J. Invest. Dermatol.* **135**, S32–S42.

[88] Kudo T, Takeda M (2001) Alzheimer disease. *Nippon Rinsho* **59 Suppl 8**, 504–510.

[89] Salmon DP, Bondi MW (2010) Neuropsychological Assessment of Dementia. *Annu Rev Psychol* **60**, 257–282.

[90] Chaves MLF, Godinho CC, Porto CS, Mansur L, Carthery-Goulart MT, Yassuda MS, Beato R (2011) Avaliação cognitiva, comportamental e funcional: Doença de Alzheimer. *Dement. e Neuropsychol.* **5**, 153–166.

[91] Harvey PD (2012) Clinical applications of neuropsychological assessment. *Dialogues Clin. Neurosci.* **14**, 91–99.

[92] Wattmo C, Minthon L, Wallin ÅK (2016) Mild versus moderate stages of Alzheimer's disease: Three-year outcomes in a routine clinical setting of cholinesterase inhibitor therapy. *Alzheimer's Res. Ther.* **8**,.

[93] Benson AD, Slavin MJ, Tran T-T, Petrella JR, Doraiswamy PM (2005) Screening for Early Alzheimer's Disease: Is There Still a Role for the Mini-Mental State Examination? *Prim. Care Companion J. Clin. Psychiatry* **7**, 62–69.

[94] Samtani MN, Raghavan N, Novak G, Nandy P, Narayan VA (2014) Disease progression model for Clinical Dementia Rating-Sum of Boxes in mild cognitive impairment and Alzheimer's subjects from the Alzheimer's Disease Neuroimaging Initiative. *Neuropsychiatr. Dis. Treat.* **10**, 929–952.

[95] Kim JW, Byun MS, Sohn BK, Yi D, Seo EH, Choe YM, Kim SG, Choi HJ, Lee JH, Chee IS, Woo JI, Lee DY (2017) Clinical dementia rating orientation score as an excellent predictor of the progression to alzheimer's disease in mild cognitive impairment. *Psychiatry Investig.* **14**, 420–426.

[96] Younesi E, Hofmann-Apitius M (2013) Biomarker-guided translation of brain imaging into disease pathway models. *Sci. Rep.* **3**,.

[97] Saeed U, Compagnone J, Aviv RI, Strafella AP, Black SE, Lang AE, Masellis M (2017) Imaging biomarkers in Parkinson's disease and Parkinsonian syndromes: Current and emerging concepts. *Transl. Neurodegener.* **6**, 1–25.

[98] Locascio JJ, Atri A (2011) An Overview of Longitudinal Data Analysis Methods for

Neurological Research. *Dement. Geriatr. Cogn. Dis. Extra* **1**, 330–357.

[99] Dickerson BC, Sperling RA (2005) Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease. *NeuroRx* **2**, 348–60.

[100] Ruan Q, D'Onofrio G, Sancarlo D, Bao Z, Greco A, Yu Z (2016) Potential neuroimaging biomarkers of pathologic brain changes in Mild Cognitive Impairment and Alzheimer's disease: A systematic review. *BMC Geriatr.* **16**, 1–9.

[101] Varghese T, Sheelakumari R, James JS, Mathuranath PS (2013) A review of neuroimaging biomarkers of Alzheimer's disease. *Neurol. Asia* **18**, 239–248.

[102] Ferreira D, Verhagen C, Hernández-Cabrera JA, Cavallin L, Guo CJ, Ekman U, Muehlboeck JS, Simmons A, Barroso J, Wahlund LO, Westman E (2017) Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: Longitudinal trajectories and clinical applications. *Sci. Rep.* **7**, 1–13.

[103] Acosta-Cabronero J, Nestor PJ (2014) Diffusion tensor imaging in Alzheimer's disease: Insights into the limbic-diencephalic network and methodological considerations. *Front. Aging Neurosci.* **6**, 1–21.

[104] Müller H-P, Kassubek J (2013) Diffusion Tensor Magnetic Resonance Imaging in the Analysis of Neurodegenerative Diseases. *J. Vis. Exp.*

[105] Oishi K, Mielke MM, Albert M, Lyketsos CG, Mori S (2011) DTI analyses and clinical applications in Alzheimer's disease. *Adv. Alzheimer's Dis.* **2**, 525–534.

[106] Kantarci K, Murray ME, Schwarz CG, Reid RI, Przybelski SA, Lesnick T, Zuk SM, Raman MR, Senjem ML, Gunter JL, Boeve BF, Knopman DS, Parisi JE, Petersen RC, Jack CR, Dickson DW (2017) White-matter integrity on DTI and the pathologic staging of Alzheimer's disease. *Neurobiol. Aging* **56**, 172–179.

[107] Yang G, Sau C, Lai W, Cichon J, Li W (2015) HHS Public Access. **344**, 1173–1178.

[108] Smith AC, Blackshaw JA, Robinson AJ (2012) MitoMiner: A data warehouse for mitochondrial proteomics data. *Nucleic Acids Res.* **40**, 1160–1167.

[109] Sierra-marcos A (2017) Regional Cerebral Blood Flow in Mild Cognitive Impairment and Alzheimer ' s Disease Measured with Arterial Spin Labeling Magnetic Resonance

Imaging. **2017**,.

[110] Galvin J, Price J, Yan Z, Morris J, Sheline Y (2011) Resting bold fMRI differentiates dementia with Lewy bodies vs. Alzheimer's disease. *Alzheimer's Dement.* **7**, e69.

[111] Vemuri P, Jones DT, Jack CR (2012) Resting state functional MRI in Alzheimer's Disease. *Alzheimers. Res. Ther.* **4**, 2.

[112] Marcus C, Mena E, Subramaniam RM (2014) Brain PET in the diagnosis of Alzheimer's disease. *Clin. Nucl. Med.* **39**, e413–e426.

[113] Chintamaneni M, Bhaskar M (2012) Biomarkers in Alzheimer's Disease: A Review. *ISRN Pharmacol.* **2012**, 1–6.

[114] DeBay DR, Reid GA, Pottie IR, Martin E, Bowen C V., Darvesh S (2017) Targeting butyrylcholinesterase for preclinical single photon emission computed tomogrpahy (SPECT) imaging of Alzheimer's disease. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **3**, 166–176.

[115] Arora A, Bhagat N (2016) Insight into the Molecular Imaging of Alzheimer's Disease. *Int. J. Biomed. Imaging* **2016**,.

[116] Weih M, Degirmenci Ü, Kreil S, Lewczuk P, Schmidt D, Kornhuber J, Kuwert T (2010) Perfusion Imaging with SPECT in the Era of Pathophysiology-Based Biomarkers for Alzheimer's Disease. *Int. J. Alzheimers. Dis.* **2010**, 1–5.

[117] Morris GP, Clark IA, Vissel B (2014) Inconsistencies and Controversies Surrounding the Amyloid Hypothesis of Alzheimer's Disease. *Acta Neuropathol. Commun.* **2**, 1–21.

[118] Wang S, Mims PN, Roman RJ, Fan F (2016) Is Beta-Amyloid Accumulation a Cause or Consequence of Alzheimer's Disease. *of* **4**, 2.

[119] Stanley ML, Moussa MN, Paolini BM, Lyday RG, Burdette JH, Laurienti PJ (2013) Defining nodes in complex brain networks. *Front. Comput. Neurosci.* **7**, 1–14.

[120] Gomez-Ramirez J, Wu J (2014) Network-based biomarkers in Alzheimer's disease: Review and future directions. *Front. Aging Neurosci.* **6**, 1–9.

[121] Johnson CH, Ivanisevic J, Siuzdak G (2016) Metabolomics: Beyond biomarkers and

towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459.

[122] Wiepert DA, Lowe VJ, Knopman DS, Boeve BF, Graff-Radford J, Petersen RC, Jack CR, Jones DT (2017) A robust biomarker of large-scale network failure in Alzheimer's disease. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **6**, 152–161.

[123] Wang H, Qiu T, Shi J, Liang J, Wang Y, Quan L, Zhang Y, Zhang Q, Tao K (2016) Gene expression profiling analysis contributes to understanding the association between non-syndromic cleft lip and palate, and cancer. *Mol. Med. Rep.* **13**, 2110–2116.

[124] Moustafa AA, Hassan M, Hewedi DH, Hewedi I, Garami JK, Al Ashwal H, Zaki N, Seo SY, Cutsuridis V, Angulo SL, Natesh JY, Herzallah MM, Frydecka D, Misiak B, Salama M, Mohamed W, El Haj M, Hornberger M (2017) Genetic underpinnings in Alzheimer's disease - A review. *Rev. Neurosci.* **29**, 21–38.

[125] Enger R, Gundersen G, Haj-Yasein N, Eilert-Olsen M, Thoren A, Vindedal G, PH P, Skare O, Nedergaard M, Nagelhus E (2012) NIH Public Access. *December* **60**, 2018–2026.

[126] Santiago JA, Bottero V, Potashkin JA (2017) Dissecting the molecular mechanisms of neurodegenerative diseases through network biology. *Front. Aging Neurosci.* **9**,.

[127] Wang J, Chen L, Wang Y, Zhang J, Liang Y, Xu D (2013) A Computational Systems Biology Study for Understanding Salt Tolerance Mechanism in Rice. *PLoS One* **8**,.

[128] Bai Y, Dong Z, Shang Q, Zhao H, Wang L, Guo C, Gao F, Zhang L, Wang Q (2016) Pdcd4 is involved in the formation of stress granule in response to oxidized low-density lipoprotein or high-fat diet. *PLoS One* **11**, 1–17.

[129] Huang W, Zhang X, Chen W (2016) Role of oxidative stress in Alzheimer's disease (Review). *Biomed. Reports* 519–522.

[130] Manuscript A, Ras-erk T (2012) NIH Public Access. **36**, 320–328.

[131] Santarpia LL, Lippman S, El-Naggar A (2012) Targeting the Mitogen-Activated Protein Kinase RAS-RAF Signaling Pathway in Cancer Therapy. *Expert Opin Ther Targets* **16**, 103–119.

[132] Hallock P, Thomas MA (2012) Integrating the Alzheimer's disease proteome and transcriptome: a comprehensive network model of a complex disease. *OMICS* **16**, 37–49.

[133] Xia J, Rocke DM, Perry G, Ray M (2015) Differential Network Analyses of Alzheimer's Disease Identify Early Events in Alzheimer's Disease Pathology. *Int. J. Alzheimers. Dis.* **2014**,.

[134] Gómez Ravetti M, Rosso OA, Berretta R, Moscato P (2010) Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gene expression profiles in Alzheimer's disease. *PLoS One* **5**,.

[135] Groups S (2015) Panels, Mini-Panels and Study GroupsMonday, December 7, 2015. *Neuropsychopharmacology* **40**, S1–S105.

[136] Haure-Mirande JV, Audrain M, Fanutza T, Kim SH, Klein WL, Glabe C, Readhead B, Dudley JT, Blitzer RD, Wang M, Zhang B, Schadt EE, Gandy S, Ehrlich ME (2017) Deficiency of TYROBP, an adapter protein for TREM2 and CR3 receptors, is neuroprotective in a mouse model of early Alzheimer's pathology. *Acta Neuropathol.* **134**, 769–788.

[137] Shen L, Jia J (2016) An Overview of Genome-Wide Association Studies in Alzheimer's Disease. *Neurosci. Bull.* **32**, 183–190.

[138] Naz M, Younesi E, Hofmann-Apitius M (2017) Systematic Analysis of GWAS Data Reveals Genomic Hotspots for Shared Mechanisms between Neurodegenerative Diseases. *J. Alzheimer's Dis. Park.* **7**,.

[139] Pettifer S, Thorne D, McDermott P, Marsh J, Villéger A, Kell DB, Attwood TK (2009) Visualising biological data: A semantic approach to tool and database integration. *BMC Bioinformatics* **10**, 1–12.

[140] Hoehndorf R, Schofield PN, Gkoutos G V. (2015) The role of ontologies in biological and biomedical research: A functional perspective. *Brief. Bioinform.* **16**, 1069–1080.

[141] Yang C, Xiao S (2015) Are the revise d diagnostic criteria for Alzheimer ' s disease useful in low- and middle-income countries ? **27**, 119–123.

[142] Malhotra A, Younesi E, G??ndel M, M??ller B, Heneka MT, Hofmann-Apitius M (2014) ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease. *Alzheimer's Dement.* **10**, 238–246.

[143] Younesi E, Malhotra A, Gündel M, Scordis P, Kodamullil AT, Page M, Müller B, Springstubbe S, Wüllner U, Scheller D, Hofmann-Apitius M (2015) PDON: Parkinson's disease ontology for representation and modeling of the Parkinson's disease knowledge domain. *Theor. Biol. Med. Model.* **12**, 20.

[144] Cook CE, Bergman MT, Finn RD, Cochrane G, Birney E, Apweiler R (2016) The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic Acids Res.* **44**, D20–D26.

[145] Brooksbank C, Cameron G, Thornton J (2005) The European Bioinformatics Institute's data resources: Towards systems biology. *Nucleic Acids Res.* **33**, 46–53.

[146] Toomula, N, Kumar A, Kumar D S, Bheemidi VS (2012) Biological Databases-Integration of Life Science Data. *J. Comput. Sci. Syst. Biol.* **4**, 87–92.

[147] Zou D, Ma L, Yu J, Zhang Z (2015) Biological databases for human research. *Genomics, Proteomics Bioinforma.* **13**, 55–63.

[148] Stoesser G, Baker W, van den Broek A, Garcia-Pastor M, Kanz C, Kulikova T, Leinonen R, Lin Q, Lombard V, Lopez R, Mancuso R, Nardone F, Stoehr P, Tuli MA, Tzouvara K, Vaughan R (2003) The EMBL nucleotide sequence database: Major new developments. *Nucleic Acids Res.* **31**, 17–22.

[149] Jimenez-Gutierrez LR, Barrios-Hernández CJ, Pedraza-Ferreira GR, Vera-Cala L, Martinez-Perez F (2016) Importance of databases of nucleic acids for bioinformatic analysis focused to genomics. *J. Phys. Conf. Ser.* **743**,.

[150] Edgar R (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210.

[151] Parkinson H (2004) ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **33**, D553–D555.

[152] Blake JA, Richardson JE, Bult CJ, Kadin JA, Eppig JT, Baldarelli RM, Beal JS, Bradt

DW, Burkart DL, Butler NE, Campbell J, Chu T, Corbani LE, Cousins S, Drabkin HJ, Dahmen D, Frazer K, Garippa DM, Goldsmith CW, Grant PL, Lennon-Pierce M, Lewis J, Lu I, Lutz CM, Maltais LJ, Mani P, McKenzie LM, Li N, Ormsby JE, Planchart A, Ramachandran S, Reed DJ, Shaw DR, Smith CL, Szauter P, Vanden Borre P, Washburn L, Winslow J (2003) MGD: The mouse genome database. *Nucleic Acids Res.* **31**, 193–195.

[153] Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012) Saccharomyces Genome Database: The genomics resource of budding yeast. *Nucleic Acids Res.* **40**, 700–705.

[154] Westbrook J, Feng Z, Chen L, Yang H, Berman HM (2003) The Protein Data Bank and structural genomics. *Nucleic Acids Res.* **31**, 489–491.

[155] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: The protein families database. *Nucleic Acids Res.* **42**, 222–230.

[156] Apweiler R (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32**, 115D–119.

[157] Knudsen M, Wiuf C (2010) The CATH database. *Hum. Genomics* **4**, 207–12.

[158] Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, O'Donnell L, Oster S, Theesfeld C, Sellam A, Stark C, Breitkreutz BJ, Dolinski K, Tyers M (2017) The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379.

[159] Mewes H-W, Hani J, Pfeiffer F, Frishman D (1998) MIPS: a database for protein sequences and complete genomes. *Nucleic Acids Res.* **26**, 33–37.

[160] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, Von Mering C (2015) STRING v10: Protein-protein interaction networks, integrated over

the tree of life. *Nucleic Acids Res.* **43**, D447–D452.

[161] Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucl. Acids Res.* **28**, 27–30.

[162] Ingenuity IPA - Ingenuity Pathway Analysis. 1–8.

[163] Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, MayMahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, Stein L (2011) Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, 691–697.

[164] Hood L, Rowen L (2013) The human genome project: Big science transforms biology and medicine. *Genome Med.* **5**, 1.

[165] Gligorijević V, Pržulj N (2015) Methods for biological data integration: perspectives and challenges. *J. R. Soc. Interface* **12**, 20150571-.

[166] Triplet T, Butler G (2014) A review of genomic data warehousing systems. *Brief. Bioinform.* **15**, 471–483.

[167] Schomburg I, Jeske L, Ulbrich M, Placzek S, Chang A, Schomburg D (2017) The BRENDA enzyme information system–From a database to an expert system. *J. Biotechnol.* **261**, 194–206.

[168] Smith RN, Aleksic J, Butano D, Carr A, Contrino S, Hu F, Lyne M, Lyne R, Kalderimis A, Rutherford K, Stepan R, Sullivan J, Wakeling M, Watkins X, Micklem G (2012) InterMine: A flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics* **28**, 3163–3165.

[169] Karp PD (2002) The MetaCyc Database. *Nucleic Acids Res.* **30**, 59–61.

[170] Fujibuchi W, Akiyama Y, Uchiyama I, Ogiwara A, Migimatsu H, Goto S, Kanehisa M (1998) DBGET/LinkDB: an integrated database retrieval system. *Pac. Symp. Biocomput.* 683–694.

[171] Langella S, Oster S, Hastings S, Siebenlist F, Phillips J, Ervin D, Permar J, Kurc T, Saltz J (2007) The Cancer Biomedical Informatics Grid (caBIG) Security

Infrastructure. *AMIA Annu. Symp. Proc.* 433–7.

[172] Xi H, Park J, Ding G, Lee YH, Li Y (2009) SysPIMP: The web-based systematical platform for identifying human disease-related mutated sequences from mass spectrometry. *Nucleic Acids Res.* **37**, 913–920.

[173] Perry RJ, Samuel VT, Petersen KF, Shulman GI, Haven N, Haven N (2015) HHS Public Access. **510**, 84–91.

[174] Apiletti D, Bruno G, Ficarra E, Baralis E (2006) Data cleaning and semantic improvement in biological databases. *J. Integr. Bioinform.* **3**, 1–11.

[175] Cannata N, Schröder M, Marangoni R, Romano P (2008) A Semantic Web for bioinformatics: Goals, tools, systems, applications. *BMC Bioinformatics* **9**, 1–6.

[176] Goebel R, Zilles S, Ringlstetter C, Dengel A, Grimnes GA (2008) What Is the Role of the Semantic Layer Cake for Guiding the Use of Knowledge Representation and Machine Learning in the Development of the Semantic Web? *AAAI Spring Symp. Symbiotic Relationships between Semant. Web Knowl. Eng.* 45–50.

[177] Waagmeester A, Kutmon M, Riutta A, Miller R, Willighagen EL, Evelo CT, Pico AR (2016) Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources. *PLoS Comput. Biol.* **12**, 1–11.

[178] Merrill E, Corlosquet S, Ciccarese P, Clark T, Das S (2014) Semantic Web repositories for genomics data using the eXframe platform. *J. Biomed. Semantics* **5**, S3.

[179] Stephens S, Lavigna D, Dilascio M, Luciano J (2006) Aggregation of bioinformatics data using Semantic Web technology. *Web Semant. Sci. Serv. Agents World Wide Web* **4**, 216–221.

[180] Post LJG, Roos M, Marshall SM, Van driel R, Breit TM (2007) A semantic web approach applied to integrative bioinformatics experimentation: A biological use case with genomics data. *Bioinformatics* **23**, 3080–3087.

[181] Mironov V, Seethappan N, Blondé W, Antezana E, Splendiani A, Kuiper M (2011) Gauging triple stores with actual biological data. *BMC Bioinformatics* **13**, S3.

[182] Splendiani A (2008) RDFScape: Semantic Web meets systems biology. *BMC Bioinformatics* **9 Suppl 4**, S6.

[183] Neththikumara N (2015) *Information retrieval from Biological Databases*.

[184] Merelli I, Pérez-Sánchez H, Gesing S, D'Agostino D (2014) Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. *Biomed Res. Int.* **2014**,.

[185] Duong TH, Nguyen HQ, Jo GS (2017) Smart Data: Where the Big Data Meets the Semantics. *Comput. Intell. Neurosci.* **2017**, 2–3.

[186] Antezana E, Blondé W, Egaña M, Rutherford A, Stevens R, De Baets B, Mironov V, Kuiper M (2009) BioGateway: a semantic systems biology tool for the life sciences. *BMC Bioinformatics* **10 Suppl 1**, S11.

[187] Paschke A, Burger A, Romano P, Marshall MS, Splendiani A (2010) Semantic web applications and tools for life sciences SWAT4LS - 2010. *CEUR Workshop Proc.* **698**, S1.

[188] Belleau F, Nolin M-A, Tourigny N, Rigault P, Morissette J (2008) Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *J. Biomed. Inform.* **41**, 706–16.

[189] Hu W, Qiu H, Huang J, Dumontier M (2017) OUP accepted manuscript. *Database* 1–13.

[190] Lam HYK, Marenco L, Clark T, Gao Y, Kinoshita J, Shepherd G, Miller P, Wu E, Wong GT, Liu N, Crasto C, Morse T, Stephens S, Cheung K-H (2007) AlzPharm: integration of neurodegeneration data using RDF. *BMC Bioinformatics* **8 Suppl 3**, S4.

[191] Sagotsky JA, Zhang L, Wang Z, Martin S, Deisboeck TS (2008) Life Sciences and the web: A new era for collaboration. *Mol. Syst. Biol.* **4**, 1–10.

[192] Gao Y, Kinoshita J, Wu E, Miller E, Lee R, Seaborne A, Cayzer S, Clark T (2006) SWAN: A distributed knowledge infrastructure for Alzheimer disease research. *Web Semant.* **4**, 222–228.

[193] Hunter PJ, Crampin EJ, Nielsen PMF (2008) Bioinformatics, multiscale modeling and the IUPS Physiome Project. *Brief. Bioinform.* **9**, 333–343.

[194] Sütterlin T, Kolb C, Dickhaus H, Jäger D, Grabe N (2013) Bridging the scales: Semantic integration of quantitative SBML in graphical multi-cellular models and simulations with EPISIM and COPASI. *Bioinformatics* **29**, 223–229.

[195] Buttler D, Coleman M, Critchlow T, Fileto R, Han W, Pu C, Rocco D, Xiong L (2002) Querying multiple bioinformatics information sources: Can Semantic Web research help? *SIGMOD Rec.* **31**, 59–64.

[196] Quan D (2007) Improving life sciences information retrieval using semantic web technology. *Brief. Bioinform.* **8**, 172–182.

[197] Wroe C (2006) Is semantic web technology ready for healthcare? *CEUR Workshop Proc.* **194**,.

[198] Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT, Maldonado JA (2009) A model-driven approach for representing clinical archetypes for Semantic Web environments. *J. Biomed. Inform.* **42**, 150–164.

[199] Pathak J, Kiefer RC, Chute CG (2012) Using semantic web technologies for cohort identification from electronic health records for clinical research. *AMIA Summits Transl. Sci. Proc.* **2012**, 10–9.

[200] Fernández-Breis JT, Maldonado JA, Marcos M, Legaz-García M del C, Moner D, Torres-Sospedra J, Esteban-Gil A, Martínez-Salvador B, Robles M (2013) Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts. *J. Am. Med. Informatics Assoc.* **20**,.

[201] Waldman SA, Terzic A (2011) Widening the path to personalized medicine. *Clin. Transl. Sci.* **4**, 392–394.

[202] Vaudano E (2013) the Innovative Medicines Initiative: a Public Private Partnership Model To Foster Drug Discovery. *Comput. Struct. Biotechnol. J.* **6**, e201303017.

[203] Granados Moreno P, Joly Y, Knoppers BM (2017) Public–Private Partnerships in Cloud-Computing Services in the Context of Genomic Research. *Front. Med.* **4**, 1–15.

[204] Bierkens M, van der Linden W, van Bochove K, Weistra W, Fijneman RJA, Azevedo R, Boiten J-W, Beliën J, Meijer GA (2015) tranSMART. *J. Clin. Bioinforma.* **5**, S9.

[205] Scheufele E, Aronzon D, Coopersmith R, McDuffie MT, Kapoor M, Uhrich CA, Avitabile JE, Liu J, Housman D, Palchuk MB (2014) tranSMART: An Open Source Knowledge Management and High Content Data Analytics Platform. *AMIA Jt Summits Transl Sci Proc* **2014**, 96–101.

[206] Auerbach RK, Chen B, Butte AJ (2013) Relating genes to function: Identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool. *Bioinformatics* **29**, 1922–1924.

[207] Stamatoyannopoulos JA, Snyder M, Hardison R, Ren B, Gingeras T, Gilbert DM, Groudine M, Bender M, Kaul R, Canfield T, Giste E, Johnson A, Zhang M, Balasundaram G, Byron R, Roach V, Sabo PJ, Sandstrom R, Stehling AS, Thurman RE, Weissman SM, Cayting P, Hariharan M, Lian J, Cheng Y, Landt SG, Ma Z, Wold BJ, Dekker J, Crawford GE, Keller CA, Wu W, Morrissey C, Kumar SA, Mishra T, Jain D, Byrska-Bishop M, Blankenberg D, Lajoiel BR, Jain G, Sanyal A, Chen KB, Denas O, Taylor J, Blobel GA, Weiss MJ, Pimkin M, Deng W, Marinov GK, Williams BA, Fisher-Aylor KI, Desalvo G, Kiralusha A, Trout D, Amrhein H, Mortazavi A, Edsall L, McCleary D, Kuan S, Shen Y, Yue F, Ye Z, Davis CA, Zaleski C, Jha S, Xue C, Dobin A, Lin W, Fastuca M, Wang H, Guigo R, Djebali S, Lagarde J, Ryba T, Sasaki T, Malladi VS, Cline MS, Kirkup VM, Learned K, Rosenbloom KR, Kent WJ, Feingold EA, Good PJ, Pazin M, Lowdon RF, Adams LB (2012) An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **13**, 1–5.

[208] Haas M, Stephenson D, Romero K, Gordon MF, Zach N, Geerts H (2016) Big data to smart data in Alzheimer's disease: Real-world examples of advanced modeling and simulation. *Alzheimer's Dement.* **12**, 1022–1030.

[209] Cerami EG, Bader GD, Gross BE, Sander C (2006) cPath: Open source software for collecting, storing, and querying biological pathways. *BMC Bioinformatics* **7**, 1–9.

[210] Clifford R. Jack Jr., MD1,*, Matt A. Bernstein, PhD1, Nick C. Fox, MD2, Paul Thompson, PhD3, Gene Alexander, PhD4, Danielle Harvey, PhD5, Bret Borowski, RTR1, Paula J. Britson, BS1, Jennifer L. Whitwell, PhD1, Chadwick Ward, BA1,

Anders M. Dale, PhD6, Joe MWW, Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DLG, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW, L Whitwell J, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DLG, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW, Whitwell JL, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DLG, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW, L Whitwell J, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DLG, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW, Clifford R. Jack Jr., MD1,*, Matt A. Bernstein, PhD1, Nick C. Fox, MD2, Paul Thompson, PhD3, Gene Alexander, PhD4, Danielle Harvey, PhD5, Bret Borowski, RTR1, Paula J. Britson, BS1, Jennifer L. Whitwell, PhD1, Chadwick Ward, BA1, Anders M. Dale, PhD6, Joe MWW (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods. *J Magn Reson Imaging* **27**, 685–691.

[211] Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR, Jagust WJ, Shaw LM, Toga AW, Trojanowski JQ, Weiner MW (2009) Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology* **74**, 201–209.

[212] Iturria-Medina Y, Sotero RC, Toussaint PJ, Mateos-Pérez JM, Evans AC, Weiner MW, Aisen P, Petersen R, Jack CR, Jagust W, Trojanowki JQ, Toga AW, Beckett L,

Green RC, Saykin AJ, Morris J, Shaw LM, Khachaturian Z, Sorensen G, Kuller L, Raichle M, Paul S, Davies P, Fillit H, Hefti F, Holtzman D, Mesulam MM, Potter W, Snyder P, Schwartz A, Montine T, Thomas RG, Donohue M, Walter S, Gessert D, Sather T, Jiminez G, Harvey D, Bernstein M, Fox N, Thompson P, Schuff N, Borowski B, Gunter J, Senjem M, Vemuri P, Jones D, Kantarci K, Ward C, Koeppe RA, Foster N, Reiman EM, Chen K, Mathis C, Landau S, Cairns NJ, Householder E, Taylor-Reinwald L, Lee V, Korecka M, Figurski M, Crawford K, Neu S, Foroud TM, Potkin S, Shen L, Faber K, Kim S, Nho K, Thal L, Buckholtz N, Albert M, Frank R, Hsiao J, Kaye J, Quinn J, Lind B, Carter R, Dolen S, Schneider LS, Pawluczyk S, Beccera M, Teodoro L, Spann BM, Brewer J, Vanderswag H, Fleisher A, Heidebrink JL, Lord JL, Mason SS, Albers CS, Knopman D, Johnson K, Doody RS, Villanueva-Meyer J, Chowdhury M, Rountree S, Dang M, Stern Y, Honig LS, Bell KL, Ances B, Carroll M, Leon S, Mintun MA, Schneider S, Oliver A, Marson D, Griffith R, Clark D, Geldmacher D, Brockington J, Roberson E, Grossman H, Mitsis E, De Toledo-Morrell L, Shah RC, Duara R, Varon D, Greig MT, Roberts P, Albert M, Onyike C, D'Agostino D, Kielb S, Galvin JE, Cerbone B, Michel CA, Rusinek H, De Leon MJ, Glodzik L, De Santi S, Doraiswamy PM, Petrella JR, Wong TZ, Arnold SE, Karlawish JH, Wolk D, Smith CD, Jicha G, Hardy P, Sinha P, Oates E, Conrad G, Lopez OL, Oakley M, Simpson DM, Porsteinsson AP, Goldstein BS, Martin K, Makino KM, Ismail MS, Brand C, Mulnard RA, Thai G, Mc-Adams-Ortiz C, Womack K, Mathews D, Quiceno M, Diaz-Arrastia R, King R, Weiner M, Martin-Cook K, DeVous M, Levey AI, Lah JJ, Cellar JS, Burns JM, Anderson HS, Swerdlow RH, Apostolova L, Tingus K, Woo E, Silverman DHS, Lu PH, Bartzokis G, Graff-Radford NR, Parfitt F, Kendall T, Johnson H, Farlow MR, Hake A, Matthews BR, Herring S, Hunt C, Van Dyck CH, Carson RE, MacAvoy MG, Chertkow H, Bergman H, Hosein C, Black S, Stefanovic B, Caldwell C, Hsiung GYR, Feldman H, Mudge B, Assaly M, Kertesz A, Rogers J, Bernick C, Munic D, Kerwin D, Mesulam MM, Lipowski K, Wu CK, Johnson N, Sadowsky C, Martinez W, Villena T, Turner RS, Johnson K, Reynolds B,

Sperling RA, Johnson KA, Marshall G, Frey M, Lane B, Rosen A, Tinklenberg J, Sabbagh MN, Belden CM, Jacobson SA, Sirrel SA, Kowall N, Killiany R, Budson AE, Norbash A, Johnson PL, Allard J, Lerner A, Ogrocki P, Hudson L, Fletcher E, Carmichael O, Olichney J, DeCarli C, Kittur S, Borrie M, Lee TY, Bartha R, Johnson S, Asthana S, Carlsson CM, Potkin SG, Preda A, Nguyen D, Tariot P, Reeder S, Bates V, Capote H, Rainka M, Scharre DW, Kataki M, Adeli A, Zimmerman EA, Celmins D, Brown AD, Pearlson GD, Blank K, Anderson K, Santulli RB, Kitzmiller TJ, Schwartz ES, Sink KM, Williamson JD, Garg P, Watkins F, Ott BR, Querfurth H, Tremont G, Salloway S, Malloy P, Correia S, Rosen HJ, Miller BL, Mintzer J, Spicer K, Bachman D, Finger E, Pasternak S, Rachinsky I, Drost D, Pomara N, Hernando R, Sarrael A, Schultz SK, Ponto LLB, Shim H, Smith KE, Relkin N, Chaing G, Raudin L, Smith A, Fargher K, Raj BA, Neylan T, Grafman J, Davis M, Morrison R, Hayes J, Finley S, Friedl K, Fleischman D, Arfanakis K, James O, Massoglia D, Fruehling JJ, Harding S, Peskind ER, Petrie EC, Li G, Yesavage JA, Taylor JL, Furst AJ (2016) Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nat. Commun.* **7**,.

[213] Doyle OM, Westman E, Marquand AF, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, Soininen H, Lovestone S, Williams SCR, Simmons A (2014) Predicting progression of Alzheimer's disease using ordinal regression. *PLoS One* **9**, 1–10.

[214] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P, Masters C, Milner A, Pike K, Rowe C, Savage G, Szoeke C, Taddei K, Villemagne V, Woodward M, Ames D, Group the AR (2009) The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int. Psychogeriatrics* **21**, 672.

[215] Hwang J, Kim CM, Jeon S, Lee JM, Hong YJ, Roh JH, Lee JH, Koh JY, Na DL (2016) Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **2**, 58–67.

[216] Dickerson BC, Brickhouse M, McGinnis S, Wolk DA (2017) Alzheimer's disease: The influence of age on clinical heterogeneity through the human brain connectome. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **6**, 122–135.

[217] Lee ES, Black RA, Harrington RD, Tarczy-Hornoch P, FACMI (2015) Characterizing Secondary Use of Clinical Data. *AMIA Summits Transl. Sci. Proc.* **2015**, 92–96.

[218] Costain G, Bassett AS (2013) Incomplete Knowledge of the Clinical Context as a Barrier to Interpreting Incidental Genetic Research Findings. *Am. J. Bioeth.* **13**, 58–60.

[219] Markram H (2013) Seven challenges for neuroscience. *Funct. Neurol.* **28**, 145–151.

[220] Bzdok D, Yeo BTT (2017) Inference in the age of big data: Future perspectives on neuroscience. *Neuroimage* **155**, 549–564.

[221] Menéndez-González M (2014) Biomarkers in neurodegenerative disorders: Translating research into clinical practice. *Front. Aging Neurosci.* **6**, 1–2.

[222] Yau WW, Tudorascu DL, Mcdade EM, James JA, Minhas D, Mowrey W, Sheu K, Snitz BE, Weissfeld L, Gianaros PJ, Aizenstein J, Price JC, Mathis CA, Lopez OL (2016) study. **14**, 804–813.

[223] Lytton WW, Neymotin SA, Kerr CC (2014) Multiscale modeling for clinical translation in neuropsychiatric disease. *J. Comput. Surg.* **1**, 7.

[224] Zanin M, Chorbev I, Stres B, Stalidzans E, Vera J, Tieri P, Castiglione F, Groen D, Zheng H, Baumbach J, Schmid JA, Basilio J, Klimek P, Debeljak N, Rozman D, Schmidt HHHW (2017) Community effort endorsing multiscale modelling, multiscale data science and multiscale computing for systems medicine. *Brief. Bioinform.* 1–6.

[225] Lytton WW, Arle J, Bobashev G, Ji S, Klassen TL, Marmarelis VZ, Schwaber J, Sherif MA, Sanger TD (2017) Multiscale modeling in the clinic: diseases of the brain and nervous system. *Brain Informatics* **4**, 219–230.

[226] Kurkinen M (2017) The amyloid hypothesis is too good to be true. *Alzheimer's, Dement. Cogn. Neurol.* **1**, 1–9.

[227] Varilly P, Chandler D (2012) NIH Public Access. **v**, 265–275.