A Knowledge-based Integrative Modeling Approach for *In-Silico* Identification of Mechanistic Targets in Neurodegeneration with Focus on Alzheimer's Disease

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Erfan Younesi

aus

Hamedan, IRAN

Bonn 2014

Angefertigt mit Genehmigung
der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinichen Friedrich-Wilhelms-Universität Bonn

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

Tag der Promotions: Dienstag, 06 Mai
Erscheinungsjahr: 2014

ABSTRACT

Dementia is the progressive decline in cognitive function due to damage or disease in the body beyond what might be expected from normal aging. Based on neuropathological and clinical criteria, dementia includes a spectrum of diseases, namely Alzheimer's dementia, Parkinson's dementia, Lewy Body disease, Alzheimer's dementia with Parkinson's, Pick's disease, Semantic dementia, and large and small vessel disease. It is thought that these disorders result from a combination of genetic and environmental risk factors.

Despite accumulating knowledge that has been gained about pathophysiological and clinical characteristics of the disease, no coherent and integrative picture of molecular mechanisms underlying neurodegeneration in Alzheimer's disease is available. Existing drugs only offer symptomatic relief to the patients and lack any efficient disease-modifying effects. The present research proposes a knowledge-based rationale towards integrative modeling of disease mechanism for identifying potential candidate targets and biomarkers in Alzheimer's disease. Integrative disease modeling is an emerging knowledge-based paradigm in translational research that exploits the power of computational methods to collect, store, integrate, model and interpret accumulated disease information across different biological scales from molecules to phenotypes. It prepares the ground for transitioning from 'descriptive' to "mechanistic" representation of disease processes.

The proposed approach was used to introduce an integrative framework, which integrates, on one hand, extracted knowledge from the literature using semantically supported text-mining

technologies and, on the other hand, primary experimental data such as gene/protein expression or imaging readouts. The aim of such a hybrid integrative modeling approach was not only to provide a consolidated systems view on the disease mechanism as a whole but also to increase specificity and sensitivity of the mechanistic model by providing disease-specific context. This approach was successfully used for correlating clinical manifestations of the disease to their corresponding molecular events and led to the identification and modeling of three important mechanistic components underlying Alzheimer's dementia, namely the CNS, the immune system and the endocrine components. These models were validated using a novel *in-silico* validation method, namely biomarker-guided pathway analysis and a pathway-based target identification approach was introduced, which resulted in the identification of the MAPK signaling pathway as a potential candidate target at the crossroad of the triad components underlying disease mechanism in Alzheimer's dementia.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my PhD supervisor, Prof. Dr. Martin Hofmann-Apitius who supported me both intellectually and professionally in all conditions over the past five years. Under his guidance, I was able to supervise 8 Master theses, write project proposals, and run industrial projects. He provided me with all degrees of freedom that I needed to grow as a scientist and I am greatly owed to him.

I want to acknowledge Prof. Dr. Holger Fröhlich for accepting to be my co-supervisor. His friendly support and constructive discussions were helpful for writing this thesis.

Special thanks go to Prof. Dr. Joachim Schulze and Prof. Dr. Evi Kostenis from the university of Bonn for accepting my invitation to be referees in the thesis jury.

I am grateful of all international Master's and PhD students working at the Bioinformatics Department, Fraunhofer SCAI for their friendly interaction and scientific support whenever I needed them. I also thank all the colleagues at Fraunhofer Institute SCAI, either in the Bioinformatics Department or in the HR department and other institutes for their scientific and administrative support. I would like to extend my gratitude to the secretariat of the Department, Meike Knieps, for helping me out with the formalities of my work.

And finally,

I dedicate this work to my intellectual parents in Iran and my lovely wife – Simin- for their patience and support during the harsh times I had in my stressful working environment.

PUBLICATION LIST

1. **Younesi, E.,** & Hofmann-Apitius, M. (2013) Biomarker-guided translation of brain imaging into disease pathway models. *Nature Scientific Reports, 3:3375.*

2. **Younesi, E.,** & Hofmann-Apitius, M. (2013). From integrative disease modeling to predictive, preventive, personalized and participatory (P4) medicine. *The EPMA journal*, 4: 23.

3. **Younesi, E.,** & Hofmann-Apitius, M. (2013) CSEO – The Cigarette Smoke Exposure Ontology. *Journal of Biomedical Semantics,* In Press.

4. **Younesi, E.**, & Hofmann-Apitius, M. (2013). A network model of genomic hormone interactions underlying dementia and its translational validation through serendipitous off-target effect. *Journal of Translational Medicine*, 11: 177.

5. Deyati, A., **Younesi, E.**, Hofmann-Apitius, M., & Novac, N. (2012). Challenges and opportunities for oncology biomarker discovery. *Drug Discovery Today*, 18: 614-624.

6. Malhotra, A., **Younesi, E.**, Gündel, M., Müller, B., Heneka, M. T., & Hofmann-Apitius, M. (2013). ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease. *Alzheimer's & Dementia*. In Press.

7. Malhotra, A., **Younesi, E.**, Gurulingappa, H., & Hofmann-Apitius, M. (2013). 'HypothesisFinder:'A Strategy for the Detection of Speculative Statements in Scientific Text. *PLoS Computational Biology*, 9(7), e1003117.

8. Gündel, M., **Younesi, E.**, Malhotra, A., Wang, J., Li, H., Zhang, B., ... & Hofmann-Apitius, M. (2013). HuPSON: the human physiology simulation ontology. *Journal of Biomedical Semantics*, 4: 35.

9. Shahid, M., Shahzad Cheema, M., Klenner, A., **Younesi, E.**, & Hofmann‐Apitius, M. (2013). SVM Based Descriptor Selection and

Classification of Neurodegenerative Disease Drugs for Pharmacological Modeling. *Molecular Informatics,* 32: 241-249.

10. **Younesi, E.**, Toldo, L., Müller, B., Friedrich, C. M., Novac, N., Scheer, A., Hofmann-Apitius, M. & Fluck, J. (2012). Mining biomarker information in biomedical literature. *BMC Medical Informatics and Decision Making*, 12: 148.

11. Zacher, B., Abnaof, K., Gade, S., **Younesi, E.**, Tresch, A., & Fröhlich, H. (2012). Joint Bayesian inference of condition-specific miRNA and transcription factor activities from combined gene and microRNA expression data. *Bioinformatics*, 28: 1714-1720.

12. Ivchenko, O., **Younesi, E.**, Shahid, M., Wolf, A., Müller, B., & Hofmann-Apitius, M. (2011). PLIO: an ontology for formal description of protein–ligand interactions. *Bioinformatics*, 27: 1684-1690.

13. Hofman-Apitius, M., **Younesi, E.**, & Kasam, V. (2009). Direct use of information extraction from scientific text for modeling and simulation in the life sciences. *Library Hi Tech*, 27: 505-519.

TABLE OF CONTENTS

LIST OF FIGURES

Note: Figures taken from various resources are presented for the academic purpose only.

LIST OF TABLES

CHAPTER 1. Introduction to Scientific Challenge

1.1. Complexity of the human brain

The human brain is the most complex organ in the human body that processes thought, action, memory, cognition and feeling. More than thousand years ago, the Persian polymath – Avicenna – whose book "Canon of Medicine" provides a complete system of medicine (known as systems medicine today; http://en.wikipedia.org/wiki/Avicenna) extensively described the brain function, brain diseases and a five-cell profile of human head including common sense, retentive imagination, and cognitive faculties such as thinking, estimation and memory, both separately and interconnected [1],[2].

That ancient five-dimensional anatomy of the human brain has been nowadays transformed into a more modern schema with seven scales (Figure 1).

Today, the complexity of the human brain can be viewed from two perspectives: anatomical complexity and functional complexity. Anatomically, brain has approximately 10^{11} neurons that communicate together via about 10^{14} synapses [3]. Surprisingly, there is no quantitative information on the number of brain structural parts in the public domain but as I report in details in Chapter 5, currently the total number of brain anatomical parts based on known

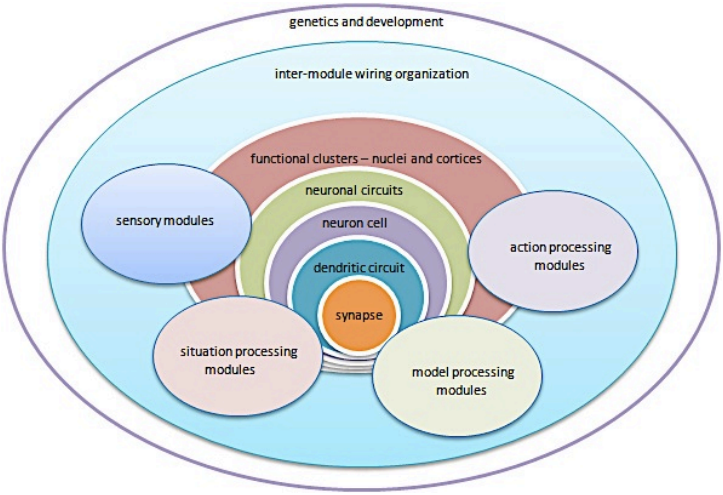


Figure 1. A schematic representation of dimensions of complexity in the functioning brain. With advancements in neurology, a more complex picture of the interconnected brain structure and its functioning appears, which poses the challenge of dealing with high-dimensional, high-order data sets at various biological levels and scales. (Taken from [http://questioneverything.typepad.com/]; last accessed: 22.10.2013).

different cell types amounts to 737. From the interconnectivity of such a complex anatomy then arises high-order functioning of the brain, which has been recently characterized based on anatomical and functional connectivity and synchronisation of their interactions [4]. Advanced neuroimaging technologies have revealed the existence of functional connectivity profiles of different brain regions, defined and visualized as complex functional networks [5]. Such a complexity is inherited by neurological diseases as well. Therefore, understanding the complexities of the brain is essential

for understanding pathologies of brain disorders, including neurodegenerative diseases.

1.2. Neurodegeneration and dementia

Since this work is focused on neurodegeneration and dementia, provisioning their definitions will be helpful for understanding the rest of the chapters in this thesis. The term neurodegeneration has been defined as "any pathological condition in which the nervous system or nerve cell loses its function, structure, or both [6]. The neuronal cell degeneration and death is quite devastating (due to limited brain's ability to replace lost neurons) and observed in the progressive neurological disorders such as Alzheimer's disease.

Dementia by definition is "the loss of intellectual functions, such as thinking and memory, which interferes with daily function" [7]. It should be noted that the term dementia does not indicate a disease in itself, but rather a group of symptoms that are dependent on a particular condition and result in cognitive decline and significant deterioration of memory.

1.3. Research motivation

Despite the tremendous advancement of the technology in contemporary times, there is little known about the complexity of the brain and its disorders resulting from neurodegeneration; hence, brain and central nervous system disorders remain the world's leading cause of disability with no proper diagnosis and very limited symptomatic treatment options. This very fact is reflected in Figure 2: while modern science has been successful in reducing the

mortality of patients suffering from cancers or cardiovascular diseases during the past decade, it has failed to control the high death rate associated with neurodegenerative diseases (NDDs) such as Alzheimer's disease (AD). Unfortunately, this high rate of mortality has been coupled with a wave of high prevalence and incidence of Alzheimer's disease, at least in the US population when compared to other NDDs. Moreover, although the worldwide prevalence of AD is estimated to quadruple by 2050, currently there exists no means to prevent, slow down, or cure AD. A very recent study indicates that, despite of reduced total mortality of people aged 55-74 years old in developed countries, total neurological deaths (due to nervous diseases and Alzheimer's) increased significantly over the period 1979-2010 [8]. Thus, the social pressure is rising as patients ask their physicians "why does research take so long? Why don't we have effective therapies for such a devastating disease?" [9].

Since 1960, many features of this fatal disease have been elucidated and many potentially disease-modifying compounds have made their way to initial human clinical trials. However, more than 200 AD drug candidates have failed to date [10]. These failures on the side of clinical application – despite significant successes on the side of molecular findings - uncover the fact that there is a deep "translational gap" between basic research and clinical application in Alzheimer's disease. Some authors have attributed these problems to the failure of the translation of research in animal studies or the use of wrong animal model [11]. But repetitive failure of amyloid-targeting clinical trials has led to growing skepticism that the hypothesis on which AD trials are based might not be valid [12].

Indeed there are important questions about pathogenic mechanism of AD to be answered; for instance, "are different pathways leading to the manifestation of AD?", "are plaques and tangles secondary events or downstream effects?", or "what role is the dimension of time playing in the pathophysiology of AD?". The current situation urges for a better understanding of the disease mechanism as it progresses over time and highlights the pressing need for correct and precise prediction of efficient mechanistic targets right at the beginning of the drug discovery pipeline.

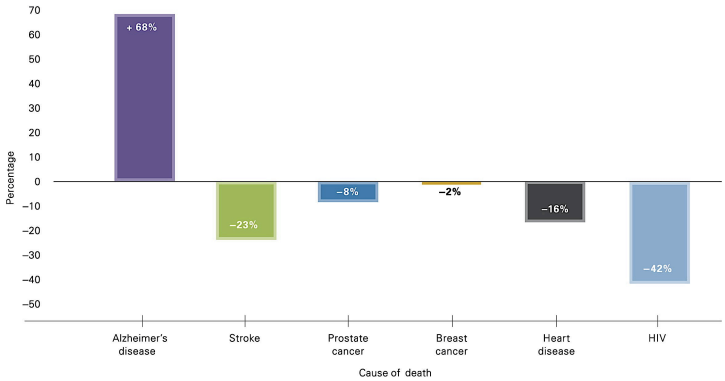


Figure 2. Changes in selected causes of death between 2000 and 2008. The figure shows the overall mortality (in the given timeframe of 8 years) in selected disease areas. Note that there is substantial controversy about speaking of a "AD mortality", as AD is not the cause of death but rather an accompanying condition.

(Taken from [http://www.alzheimersanddementia.com /article/ S15525260(13)00076-9/abstract]; last accessed: 22.10.2013).

Currently no coherent and integrative picture of molecular mechanisms underlying neurodegeneration in Alzheimer's disease is available. The AD research community is now beginning to realize that AD is a complex multifactorial disease and other non-amyloid targets, pathways, and processes should be investigated to identify novel treatment strategies [13]. Existing drugs only offer symptomatic relief to the patients and lack any efficient disease-modifying effects [14].

1.4. The goal of this thesis

The goal of the research presented in this thesis is:

a. to propose an integrative knowledge- and data-driven approach for identification of novel target candidates by addressing the problem of complexity underlying the neurodegenerative mechanisms in Alzheimer's disease.

b. to demonstrate that knowledge-based computational models and maps of Alzheimer's disease progression enable the identification of mechanistic pathways and improve the understanding of the correlations between genotypes and phenotypes at the molecular level.

CHAPTER 2. Complex Biology Underlying Neuro-degeneration in Alzheimer's Disease

2.1. Classification of neurodegenerative disorders (NDDs)

Neurodegeneration represents a large group of heterogeneous disorders in which the neuronal damage and death occurs within specific anatomical and functional areas of the brain. The exact number of NDDs is not clear but it is estimated to be a few hundred [15]. Due to the inherent anatomical and functional complexity of the brain, classification of NDDs is a challenging and complicated task, which has so far remained controversial. This is because many NDDs show clinical and pathological overlap with one another and several regions of the brain are affected simultaneously. For example, multiple system atrophy (MSA) in which several areas of brain are affected by degeneration of neurons, parkinsonism is a prominent feature but it may be accompanied with other symptoms such as sever ataxia or autonomic failure, depending on the affected area of the brain. Another example is the clinical overlap between Parkinson's disease with dementia and Alzheimer's disease dementia. Since classification of these two neurodegenerative diseases is merely based on their clinical signs, early diagnosis of Parkinson's disease (PD) is hampered. The extent of clinical overlap between these two conditions is so much greater than chance that

even some authors propose that AD and PD belong to a spectrum of neurodegenerative disorders with common disease mechanism but triggered by different etiological factors [16]. Moreover, even many diseases with neurodegeneration as their major element such as epilepsy, multiple sclerosis and schizophrenia are not classified as degenerative. To date, categorization of neurodegenerative disorders is still based on clinical features (symptomatology) or the topography of lesions (as revealed by imaging techniques). Accordingly, Przedborski and coworkers (2003) have proposed a high-level classification for NDDs (Figure 3).

Traditionally, the current nosology of NDDs follows the "one disease, one pathology" paradigm which is based on the notion of a discrete and clear correlation between disease states and certain pathological markers. For instance, early findings in autopsies of AD patients showed the presence of amyloid plaques and neurofibrillary tangles and similarly, abnormal protein aggregates of Lewy bodies were identified in the postmortem brains of Parkinson's patients. However, this kind of nosology poses several problems.

Firstly, the progressive nature of NDDs highlights the point that neurological changes are "time-dependent" and for this reason, it is difficult to distinguish disease stages or decide about the associated neuropathology. Due to this long "latency period" in NDDs generally, and in dementias particularly, which can take several decades as long as 22 years before the diagnosis of dementia, clinically it is not possible to operationalize poor cognitive performance in elderly [17]. Secondly, problems relate to labeling the disease when more than one possible diagnosis exists for a set of

symptoms. Disease labeling may change over time as the disease progresses and this can be misleading for patients [18].

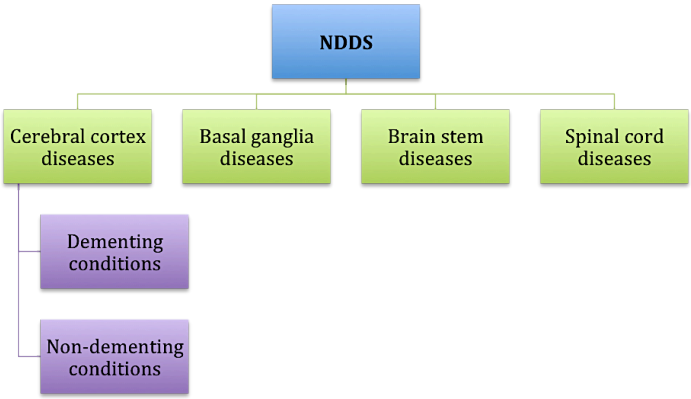


Figure 3. High-level classification schema for grouping NDDs based on anatomical structure. This schema reflects the conventional anatomically guided classification system, which does not take into consideration the disease mechanism. Reproduced from [15].

As mentioned before, the overlapping symptomatology of NDDs suggests that single pathological markers can poorly correlate with clinical symptoms and perhaps, multiple markers or molecular features can better represent separate NDD etiologies. Recent initiatives have been launched to develop a new taxonomy of diseases based on their underlying molecular and environmental causes rather than on physical signs and symptoms [19],[20]. Therefore, NDDs should be seen as a spectrum of diseases with overlapping symptoms and mixed pathologies.

2.2. Dementias and Alzheimer's disease

2.2.1. Nosology

Today, several classification systems for dementia subtypes exist, taking into account the etiological specifications, anatomic location, course of disease and prognosis (http://www.ninds.nih.gov/disorders/dementias/detail_dementia.htm). Figure 4 summarizes all three classifications in a single schema with detailed categorization based on subcortical subtypes. This schema integrates the entire concept of "clinical vs. histopathological vs. mechanism-based classification". The limitation of this classification schema is that it does not provide sufficient granularity at the etiological level.

In contrast, an etiology-based classification schema provides a better granularity and more complete list of dementia subtypes. Table 1 depicts this classification system [22]. However, this classification system is heavily biased towards anatomical etiology of dementias and does not include the knowledge of molecular basis underlying dementias.

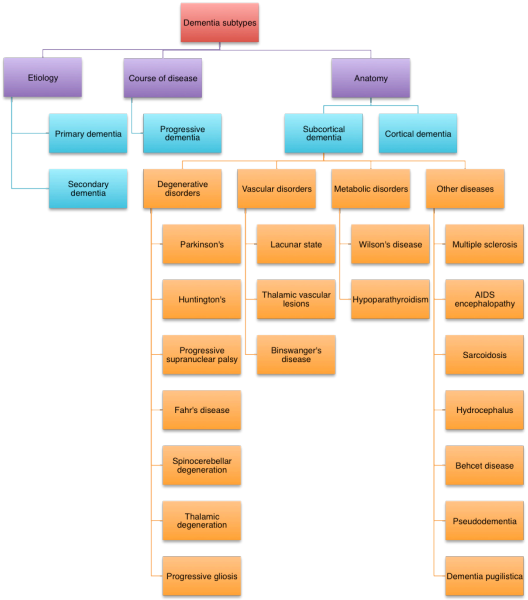


Figure 4. An integrated classification system for grouping dementia subtypes. Reproduced from [21].

The lack of proper classification of diseases with high resolution at the molecular and mechanism level has led to the ignorance of disease heterogeneity. Unfortunately, disease heterogeneity is not considered in the design and conduct of clinical trials and consequently, the likelihood of success in clinical trials for an effective drug will probably reduced. For this reason, Kola and Bell (2011) have called for reformation of the human disease taxonomy by moving away from traditional diagnostic-based criteria to molecular-based stratification of patients with multiple disease subtype [23].

Table 1. Etiology-based classification system for dementia subtypes based on anatomy

Etiology-based dementias	Primary cortical	Alzheimer's disease	
		Frontotemporal dementia	
		Pick's disease	
		Primary progressive aphasia	
		Agryophilic grain disease	
		Lewy body dementia	Diffuse
			Mixed
			Cerebral
	Primary subcortical	Parkinson's disease	
		Corticobasal degeneration	
		Progressive supranuclear palsy	
		Multiple system atrophy	
		Lewy body dementia	Transitional
			Brainstem
	Cerebrovascular	Vascular dementia	Large/small vessel strokes
			Multiple lacunar infarcts
			Binswanger disease
			CADASIL
			Cerebral amyloid angiopathy
	Structural/traumatic injury	Brain tumors	
		Limbic encephalitis	
		Traumatic brain injury	
		Dementia pugillistica	
		Chronic subdural hematoma	
		Normal pressure hydrocephalus	
		Postanoxic state	
		Postoperative cognitive dysfunction	
	Toxic exposure	Substance induced dementia	
		Medication induced dementia	
		Alcohol dementia	
		Inhalant induced dementia	
		Wernicke-Korsakoff syndrome	
		Toxic metal exposure	
		Wilson's disease	
		Toxic gas exposure	
	Nutritional deficiency	Vitamin B12 deficiency	
		Folate deficiency	

		Niacin deficiency	
		Thiamine deficiency	
	Infectious disease	Bacterial	Bacterial manengitis
			Whippie disease
		Viral	Viral manengitis
			Herpes simplex encephalitis
			HIV-associated dementia
			Progressive Leukoencephalopathy
			Sleeping sickness
			Neurosyphillis
			Lyme disease
		Fungal	Fungal manengitis
			Cryptococcal meningitis
		Prion disease	Creutzfeldt-jakob disease
			Kuru
			GSS syndrome
			Fatal familial insomnia
		Parasitic diseases	
	Organ failure	Uremic encephalopathy	
		Hepatic encephalopathy	
	Endocrine disease	Diabetes mellitus	
		Hypothyroidism	
		Hyperparathyroidism	
		Cushing syndrome	
		Addison disease	
	Neurologic/ metabolic disorders	Huntington's disease	
		Multiple sclerosis	
		Marchiafava-Bignami disease	
		Ataxia syndrome	
		Inherited storage diseases	Adrenoleukodystrophy
			Metachromatic leukodystrophy
			Cerebrotendinous xanthomonasis
	Inflammatory disease	Collagen vascular diseases	Behcet syndrome
			Sjörgen syndrome
			Systemic lupus erythematosus
		Vasculitides	Granulomatous angiitis
			Lymphomatoid granulomatosis
			Polyarteritis nodosa
		Wegener granulomatosis	

2.2.2. Epidemiology of dementias and AD

Epidemiological estimates indicate that the prevalence of AD is increasing worldwide. According to the World Health Organization in 2001, about 37 million people worldwide had been afflicted by dementia out of which 18 million suffered from AD by that time. The latest numbers come from the Alzheimer World Report, which estimated 35.6 million people with dementia in 2010, the numbers nearly doubling every 20 years, to 65.7 million in 2030 and 115.4 million in 2050 [24]. The global burden of annual incidence of dementia is shown in Figure 5. It is noteworthy that in most studies the term 'dementia' has been used interchangeably with 'AD'.

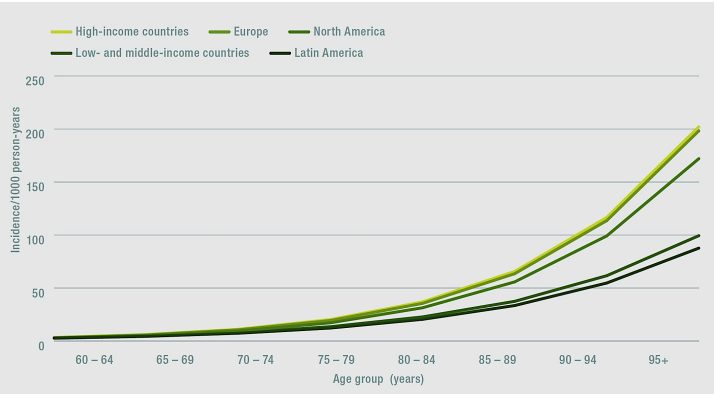


Figure 5. The global burden of annual incidence of dementia. (Taken from [http://www.who.int/mental_health/publications/dementia_report_2012/en/]; last accessed: 01.11.2013).

Epidemiological studies report that the incidence rate of new cases increases with age, almost doubling with every five-year increase in age (Figure 6).

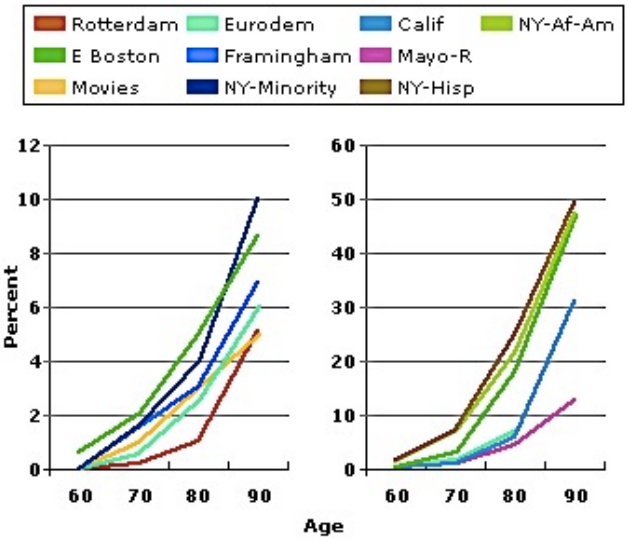


Figure 6. Projection of the incidence rate of dementia cases from 10 studies. As shown in a number of studies, the prevalence of Alzheimer's disease doubles with every five-year increase in age. (Taken from [http://www2f.biglobe.ne.jp/~boke/ adupdatecu.htm]; last accessed: 22.10.2013)

Interestingly, the incidence of dementia is lower in men but women enjoy a longer survival time than men for both AD and vascular dementia [25].

2.2.3. Risk factors for AD

Increased risk of AD has been associated with a large number of factors. The main risk factor for AD is advancing age, given that it is driven by lifelong accumulation of molecular damage. Mayoux and Stern have recently analysed the AD incidence data published in 24 studies and estimated the age-specific incidence of AD as illustrated in Figure 7 [26]. The age of onset in most AD cases is usually 65 and above, which is considered as "sporadic" but onset before this age suggests involvement of a strong genetic component. Among genetic factors, the E4 allele of the APOE gene is the only confirmed risk factor for AD, which has been shown to be specific for up to 90 percent of dementia cases [27].

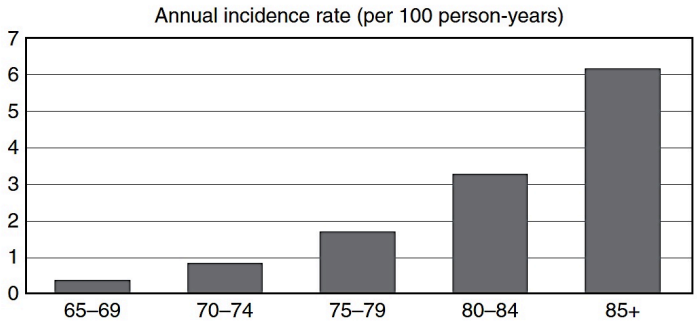


Figure 7. The annual incidence rate (per 100 person-years) for Alzheimer disease. This graph is an estimate of the data collected in 24 published studies.

Accumulating epidemiological evidence indicates that various risk factors contribute to exacerbation of neurodegenerative diseases

including gender, poor education, endocrine conditions, oxidative stress, inflammation, stroke, hypertension, diabetes, smoking, head trauma, depression, infection, tumors, vitamin deficiencies, immune and metabolic conditions, and chemical exposure [28]. Such risk factors lead to manifestation of co-morbid conditions in patients with neurodegenerative diseases. For instance, it is observed that dementia incidence is modulated by factors linked to treatment of other diseases such as diabetes [29]. Similarly, retrospective studies of clinical and autopsy data revealed that patients with rheumatoid arthritis exhibit a reduced prevalence of AD [30]. Reports also suggest that patients who take anti-inflammatory drugs or suffer from inflammatory diseases like arthritis have reduced risk of developing AD. In general, AD is associated with several major co-morbidities including hypo- and hyper-thyroidism, sleep apnea, osteoporosis, glaucoma, and rheumatoid arthritis [31]. Hence, identification of those co-morbidities that have strong etiological link to the disease may open up new intervention strategies to indirectly modify the trajectory of the disease.

Other major risk factors that modify the risk of AD include hypertension, myocardial infarction, coronary heart disease, diabetes, atherosclerosis, smoking, high cholesterol concentrations and a history of stroke (Table 2) [26]. Of the above-mentioned risk factors, age and gender and genotype can not be changed but others can be modified to prevent the disease.

Table 2. List of major factors that modify the risk of Alzheimer's dementia and their probable mechanism of effect. Adopted from [26].

Antecedent	Direction	Possible mechanisms
Cardiovascular disease	Increased	Parenchymal destruction Strategic location ↑ Aβ deposition
Smoking	Increased	Cerebrovascular effects Oxidative stress
Hypertension	Increased and decreased	Microvascular disease
Type II diabetes	Increased	Cerebrovascular effect Insulin and Aβ compete for clearance
Obesity	Increased	Increased risk of type II diabetes inflammatory
Traumatic head injury	Increased	↑Aβ and amyloid precursor protein deposition
Education	Decreased	Provides cognitive reserve
Leisure activity	Decreased	Improves lipid metabolism, mental stimulation
Mediterranean diet	Decreased	Antioxidant, anti-inflammatory
Physical activity	Decreased	Activates brain plasticity, promotes brain vascularization

2.2.4. Disease course of AD

Since AD is a progressive disease, it typically takes 10 years from diagnosis to death [32]. Some people on a course of healthy aging show only age-related changes in cognition but some other gradually move to mild cognitive impairment (MCI), which is the beginning of transition to AD. Figure 8 illustrates the natural history of AD based on clinical cognition tests.

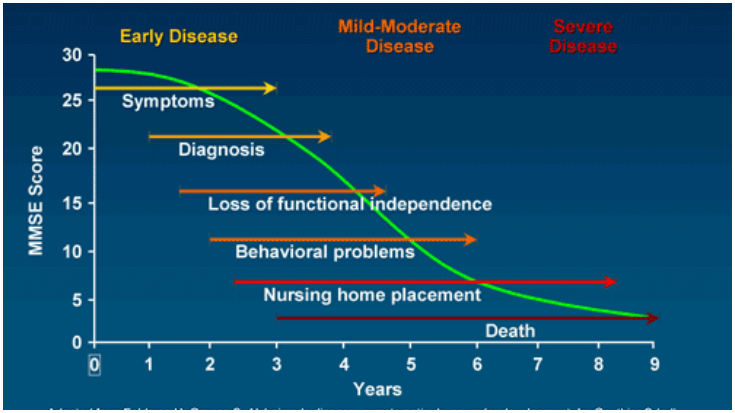


Figure 8. Natural history of Alzheimer's dementia progression based on declining cognitive function over time. All AD patients decline similarly through the stages of cognitive and functional loss as well as the stage of behavioral problems, although at different rates. (Taken from [http://www.medscape.org/viewarticle/456034_2]; last accessed: 22.10.2013)

2.2.5. Diagnosis of AD

Clinically, AD is diagnosed based on the presence of characteristic cognitive, behavioral, and neuropsychiatric symptoms such as problems with memory or speech. For research purposes, the diagnosis of AD is performed using diagnostic criteria developed and published by the DSM-IV-TR and NINCDS-ADRDA working group [33]. Given the insufficient diagnostic specificity of conventional criteria against other dementia subtypes (23-88%), this framework offers revised AD and non-AD diagnostic criteria using not only "core diagnostic criteria" but also "supportive features". The core

diagnostic criteria takes into consideration features of early and significant episodic memory impairment whereas supportive features include test for medial temporal lobe atrophy, abnormal concentrations of cerebrospinal fluid (CSF) biomarkers, specific patterns on functional neuroimaging with PET and check for autosomal dominant mutations.

CSF biomarkers are not routinely used for evaluation of AD and PET biomarkers are extensively being developed to be used as preclinical diagnostic markers. To date, there is no molecular and imaging biomarkers that could predict onset or progression of AD. However, the most popular diagnostic tests in clinic for AD are obtaining history of the onset of the patient's symptoms and their progression, physical examination including visual signs, speech, reflexes, eye movement, etc., laboratory studies such as blood test, radiological studies using 'functional' and 'structural' imaging, and finally neuropsychological tests including the most commonly used tests namely Mini-mental State Examination (MMSE) and the Mini-Cog. These tests assess mental skills of patient using several items such as recall, time and place orientation, writing skills and mental calculations.

2.2.6. Biomarkers of AD

Over the past decades, validated and disease-specific biomarkers have been introduced to explain the neuropathology of AD. These biomarkers play an important role in supporting the challenging task of definite diagnosis of AD and can be grouped in two categories: pathophysiological and topographical markers [34].

Pathophysiological biomarkers include CSF measures of reduced amyloid-beta, increased total tau, increased phosphorylated tau, and amyloid PET scanning. In contrast, topographical image-based biomarkers are less specific but identify the regional patterns of AD pathology and include medial temporal lobe atrophy and reduced glucose metabolism in temporal lobe regions. Interestingly, these imaging biomarkers have been shown to accurately map to the Braak stages of neurofibrillary tangle deposition [35].

In 2010, a hypothetical model of the major biomarkers of AD was published in which the temporal evolution of AD biomarkers in relation to each other and to the onset and progression of clinical symptoms was described. The authors have revised their model using the latest biomarker data and proposed that the two major proteinopathies underlying AD biomarker changes, Aß and tau, may be initiated independently in sporadic AD (Figure 9) [36]. In this proposed integrative model, subcortical tauopathy is the first AD pathophysiological process to arise in many individuals and is detectable only by immunostaining methods. This tauopathy, however, does not by itself lead to AD. Amyloid-beta pathophysiology arises later and independently from pre-existing tauopathy. Acceleration of the initial slowly developing subcortical tauopathy occurs after concentrations of Amyloid-beta biomarkers become abnormal (biomarker abnormality in Figure 9). Changes in imaging biomarkers then occur, followed last by onset of overt clinical symptoms.

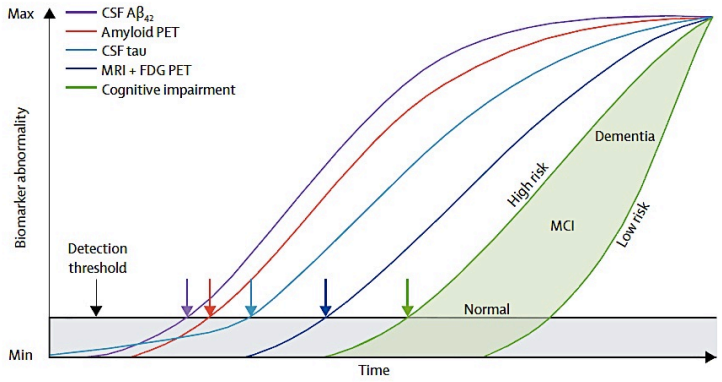


Figure 9. Model integrating tauopathy and amyloid-beta pathophysiology. The threshold for biomarker detection of pathophysiological changes is denoted by the black horizontal line. The grey area denotes the zone in which abnormal pathophysiological changes lie below the biomarker detection threshold. In this figure, tau pathology precedes A-beta deposition in time. A-beta deposition then occurs independently and rises above the biomarker detection threshold (purple and red arrows). This induces acceleration of tauopathy and CSF tau then rises above the detection threshold (light blue arrow). Later still, FDG PET and MRI (dark blue arrow) rise above the detection threshold. Finally, cognitive impairment becomes evident (green arrow), with a range of cognitive responses that depend on the individual's risk profile (light green-filled area). (Adopted from [36])

2.2.7. Stages of AD

There are different descriptions for staging in AD. Perhaps the Clinical Dementia Rating (CDR) system was the first staging scale for dementia that was developed by Hughes et al. in 1982. CDR is a five-point scale as shown in Table 3.

Table 3. Representation of the CDR scale for rating cognitive performance

CDR-0	No dementia
CDR-0.5 = Very mild	Memory problems are slight but consistent; some difficulties with time and problem solving; daily life slightly impaired.
CDR-1 = Mild	Memory loss moderate, especially for recent events, and interferes with daily activities. Moderate difficulty with solving problems; cannot function independently at community affairs; difficulty with daily activities and hobbies, especially complex ones.
CDR-2 = Moderate	More profound memory loss, only retaining highly learned material; disoriented with respect to time and place; lacking good judgment and difficulty handling problems; little or no independent function at home; can only do simple chores and has few interests.
CDR-3 = Severe	Severe memory loss; not oriented with respect to time or place; no judgment or problem solving abilities; cannot participate in community affairs outside the home; requires help with all tasks of daily living.

In 1987, Reisberg introduced the Functional Assessment Staging (FAST) system with the focus on levels of functioning and daily living activities versus cognitive decline [37]. Table 4 summarizes seven stages of the FAST system.

Table 4. Representation of the FAST system for measurement of cognitive performance

Stage 1: Normal adult	No functional decline
Stage 2: Normal older adult	Personal awareness of some functional decline.
Stage 3: Early Alzheimer's disease	Noticeable deficits in demanding job situations.
Stage 4: Mild Alzheimer's	Requires assistance in complicated tasks such as handling finances, planning parties, etc.
Stage 5: Moderate Alzheimer's	Requires assistance in choosing proper attire.
Stage 6: Moderately severe Alzheimer's	Requires assistance dressing, bathing, and toileting. Experiences urinary and fecal incontinence.
Stage 7: Severe Alzheimer's	Speech ability declines to about a half-dozen intelligible words. Progressive loss of abilities to walk, sit up, smile, and hold head up.

The Global Deterioration Scale/Functional Assessment Staging (GDS/FAST) system, developed by Auer and Reisberg in 1997 (also known as Reisberg scale), is widely used to characterize the stages of AD on a seven-stage granularity basis (Table 5) [38].

Table 5. Details of the Reisberg rating scale

Diagnosis	Stage	Signs & Symptoms
No Dementia	**Stage 1: No Cognitive Decline**	In this stage the person functions normally, has no memory loss, and is mentally healthy.
No Dementia	**Stage 2: Very Mild Cognitive Decline**	This stage is used to describe normal forgetfulness associated with aging; for example, forgetfulness of names and where familiar objects were left.
No Dementia	**Stage 3: Mild Cognitive Decline**	This stage includes increased forgetfulness, slight difficulty concentrating, decreased work performance. People may get lost more often or have difficulty finding the right words. Average duration: 7 years before onset of dementia
Early stage	**Stage 4: Moderate Cognitive Decline**	This stage includes difficulty concentrating, decreased memory of recent events, and difficulties managing finances or traveling alone to new locations. People have trouble completing complex tasks efficiently or accurately and may be in denial about their symptoms. Average duration: 2 years
Mid-stage	**Stage 5: Moderately Severe Cognitive Decline**	People in this stage have major memory deficiencies and need some assistance to complete their daily activities (dressing, bathing, preparing meals). Memory loss is more prominent and may include

		major relevant aspects of current lives. Average duration: 1.5 years
Mid-stage	**Stage 6: Severe Cognitive Decline (Middle Dementia)**	People in Stage 6 require extensive assistance to carry out daily activities. They start to forget names of close family members and have little memory of recent events. Many people can remember only some details of earlier life. They also have difficulty counting down from 10 and finishing tasks. Incontinence (loss of bladder or bowel control) is a problem in this stage. Ability to speak declines. Personality changes, such as delusions (believing something to be true that is not), compulsions (repeating a simple behavior, such as cleaning), or anxiety and agitation may occur. Average duration: 2.5 years
Late stage	**Stage 7: Very Severe Cognitive Decline (Late Dementia)**	People in this stage have essentially no ability to speak or communicate. They require assistance with most activities (e.g., using the toilet, eating). They often lose psychomotor skills, for example, the ability to walk. Average duration: 2.5 years

Reisberg and coworkers (2008) have recently provided a comparative overview of the above-mentioned staging scales in

which the typical time course of normal brain aging, MCI, and AD have been mapped to the range of stages in different scales [39].

It is worth of note that a different staging system based on the distribution pattern and packing density of amyloid deposits in autopsied brains (i.e. neuropathological staging) was proposed by Braak and Braak in 1991, which distinguished six stages [40]. This system was able to differentiate initial, intermediate, and late phases of the disease in both symptomatic and non-symptomatic individuals. This staging schema was revised in 2006 by Braak et al. using more advanced techniques with focus on tau hyperphosphorylation, as summarized in Table 6. [41].

Table 6. Updated schema for diagnosis of stages in Alzheimer's disease

Stage I	Lesions develop in the transentorhinal region
Stage II	Lesions extend into the entorhinal region
Stage III	Lesions extend into the neocortex of the fusiform and lingual gyri
Stage IV	The disease process progress more widely into neocortical association areas
Stage V	The neocortical pathology extends fanlike in frontal, superolateral, and occipital directions, and reaches the peristrate region
Stage VI	The pathology reaches the secondary and primary neocortical areas and, in the occipital lobe, extends into the striate area

Given the importance of predictive measures for identification of individuals at the risk of developing mild cognitive impairment,

characterization of the long preclinical stage of AD opens a crucial window of opportunity to intervene with therapy; to this end, recently preclinical stages of AD have been described and recommended by the National Institute on Aging (Figure 10) [42].

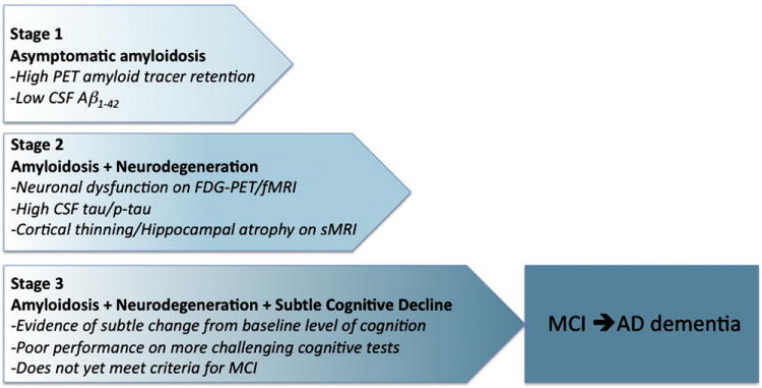


Figure 10. Graphic representation of the proposed staging framework for preclinical AD. Note that some individuals will not progress beyond Stage 1 or Stage 2. Individuals in Stage 3 are postulated to be more likely to progress to MCI and AD dementia. Abbreviations: Ab, amyloid beta; PET, position emission tomography; CSF, cerebrospinal fluid; FDG, fluorodeoxyglucose, fMRI, functional magnetic resonance imaging, sMRI, structural magnetic resonance imaging. (Adopted from [42])

The emerging concept of preclinical AD has great implications for identification of predictive biomarkers as well as development of disease-modifying treatments. For instance, it is critical to define a biomarker that best predicts progression from the preclinical to the clinical stages of AD dementia. The long latency period of AD

within the preclinical phase opens up a new opportunity for potential intervention with disease-modifying therapies. However, more studies should be conducted to determine the robustness of the above model in terms of the number of individuals transitioning from pre-clinical to clinical phase of AD. Besides, robust predictive biomarkers and preventive disease-modifying treatments are still missing and deserve more investigation.

2.3. Biology/Etiology of AD

2.3.1. Anatomical patterns of pathology in AD

Decline in cognitive abilities, learning difficulties and memory impairment, particularly the loss of episodic memory, are hallmarks of early AD. The first pre-requisite for understanding the pathology of AD is to understand the neuroanatomy of human memory and cognition because such functional tasks are crucially dependent on networks of brain areas and regional connectivity circuits. The primary brain region damaged during the course of AD is “limbic system”, which consists of several structures, most importantly, hippocampus, cingulate gyrus, limbic cortex, amygdala, olfactory bulbs, fornix, mamillary body, septum and habenular commisure (Figure 11).

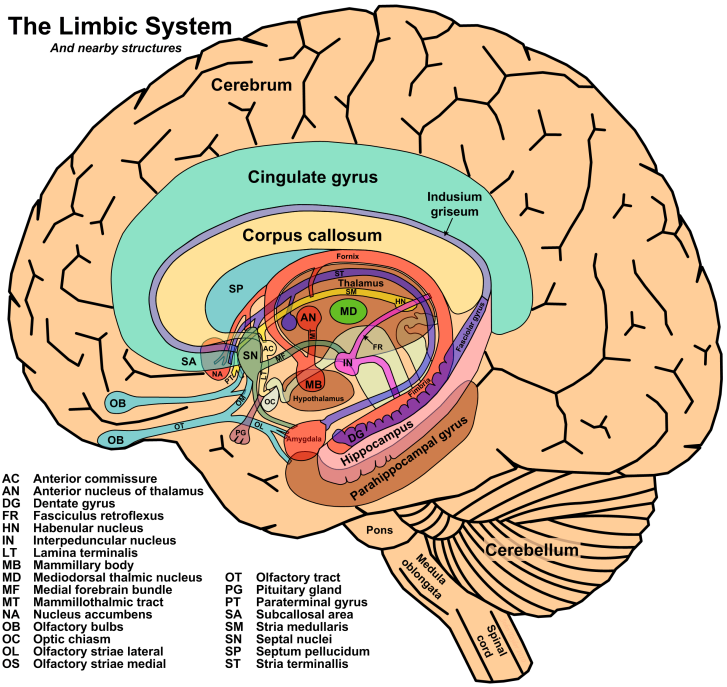


Figure 11. Neuroanatomy of the limbic system and surrounding structures in the human brain. (Taken from [http://jobspapa.com/diagram-the-limbic-system-and-surrounding-structures.html]; last accessed: 22.10.2013)

The pathology associated with AD begins in the entorhinal cortex where a decline in its volume is observed. This region plays an important role in relaying information between the hippocampus and neighboring cortices. The atrophy then spreads to hippocampus and moves out to the temporal and parietal cortices. Degeneration of entorhinal cortex occurs progressively during aging but its

pathological conversion to MCI and to AD begins long before manifestation of clinical symptoms [43].

2.3.1.1. Neuroanatomy of human memory and learning

It is long known that the medial temporal lobe is involved in processing memory. The hippocampus is a central structure for memory function and plays an important role in both spatial and episodic memories [44]. In coronal section, the hippocampus looks like an S-shaped structure and is composed of the *cornu ammonis* and the *dentate gyrus*. The *cornu ammonis* is further divided into four fields, namely CA1 to CA4 (Figure 12) [45].

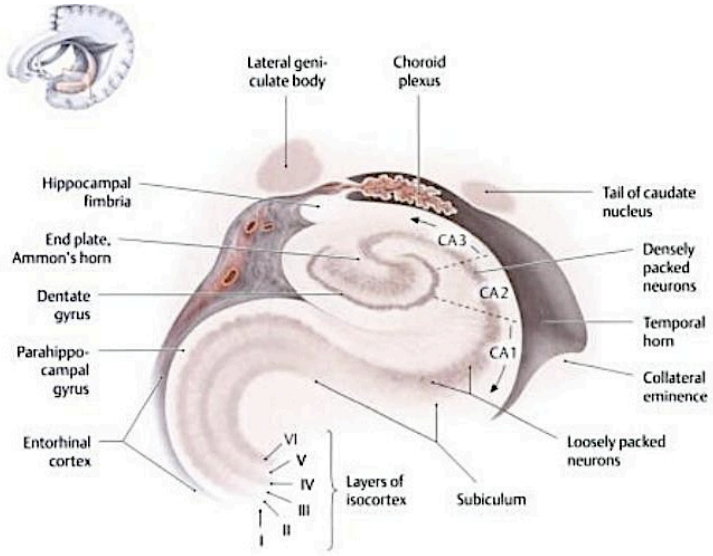


Figure 12. Neuroanatomy of the hippocampus, including CA layers.

As shown in the figure above, the hippocampus contains three layers of densely packed neurons designated CA1, CA2, and CA3. These layers can be distinguished based on differences in the density of pyramidal cells. CA1 region (also called the Sommer sector) is important in neuropathology because the first signs of neuronal death due to cerebral hypoxia are observed in this area. Dentate gyrus mainly consists of granule cells and is primarily affected during normal aging as well as very late phase of AD [46].

On a cellular network level, the hippocampus has several important connections to other brain regions through neural pathways (Figure 13). The entorhinal cortex is considered the gateway to the hippocampus because neurons forming the so called "perforant pathway" collect input from many other brain regions and deliver to the dentate gyrus of the hippocampus. "Mossy fibers" connect granule cells of the dentate gyrus to pyramidal cells of the CA3 area. The input is finally sent to the CA1 area for further processing. CA1 pyramidal cells provide output via subiculum to the entorhinal cortex and prefrontal cortex. It has been shown that the dentate gyrus and the CA3 networks are involved in pattern recognition whereas the CA1 network is involved in learning the environment. The CA1 region is particularly vulnerable to pathological effects of AD and such vulnerability has direct impact on the hippocampal-neocortical memory system [47].

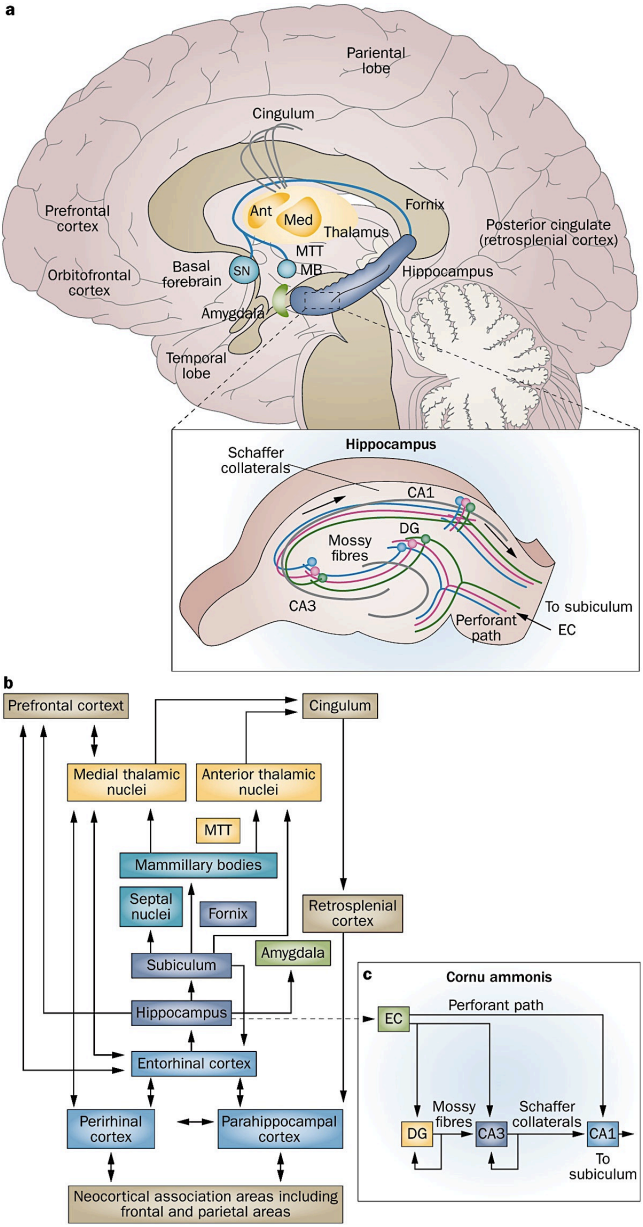

a
Pariental lobe
Cingulum
Prefrontal cortex
Ant
Med
Thalamus
Fornix
Posterior cingulate (retrosplenial cortex)
MTT
Basal forebrain
SN
MB
Hippocampus
Orbitofrontal cortex
Amygdala
Temporal lobe
Hippocampus
Schaffer collaterals
CA1
DG
Mossy fibres
CA3
To subiculum
Perforant path
EC
b
Prefrontal cortext
Cingulum
Medial thalamic nuclei
Anterior thalamic nuclei
MTT
Mammillary bodies
Septal nuclei
Fornix
Retrosplenial cortex
Amygdala
Subiculum
Hippocampus
Entorhinal cortex
Perirhinal cortex
Parahippocampal cortex
Neocortical association areas including frontal and parietal areas
c
Cornu ammonis
Perforant path
EC
Mossy fibres
DG
CA3
Schaffer collaterals
CA1
To subiculum

Figure 13. Schematic representation of the neural connections between brain regions. a) Interconnection of anatomical regions that support episodic memory, highlighting the connections between the hippocampal system, diencephalic structures and neocortical regions. b) Diagrammatic representation of the unidirectional and bidirectional network connectivity of critical structures of the temporal lobe system, diencephalic nuclei and neocortical association areas involved in memory processing. c) The anatomy of the CA, depicting the trisynaptic pathway as the principal feedforward neural circuit involved in the processing of information through the hippocampus. Abbreviations: Ant, anterior thalamic nuclei; CA, cornu ammonis; DG, dentate gyrus; EC, entorhinal cortex; MB, mamillary bodies; Med, medial thalamic nuclei; MTT, mamillothalamic tract (bundle of Vicq d'Azyr); SN, septal nuclei. (Taken from [47])

Classically, memory systems of human brain are divided into long-term and short-term (working) memory. Figure 14 illustrates further classification of human memory systems mapped to corresponding anatomical brain regions. In AD, both episodic and semantic memories are impaired. The role of semantic memory in episodic future thinking has been recently reviewed by Bartsch and Butler [48].

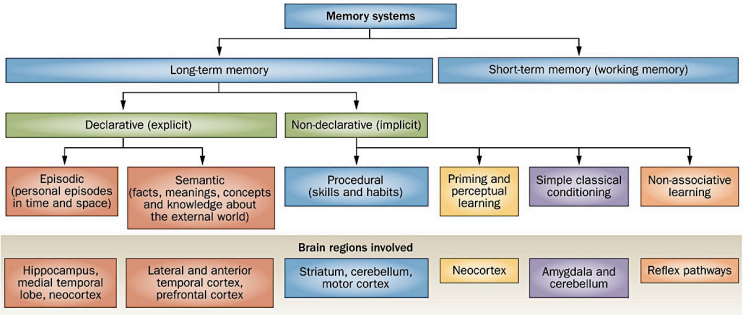


Figure 14. The classic view of memory systems based on a temporal scale. Episodic memory represents memory of specific events in time and space. Episodic autobiographical memory is usually regarded as part of episodic memory although it also comprises semantic memory, which includes facts, meanings, concepts and knowledge about the external world and refers to general factual knowledge. The different memory systems can interact and overlap with each other on a temporal scale (for example, episodic memory transfers into semantic memory over time) and content-wise (procedural learning can be modulated by declarative memory). (Taken from [47]).

2.3.1.2. Neuroanatomy of human cognition

Cognitive functions can be divided into two broad groups: basic and higher-level cognitive functions [49].

Basic cognitive functions include “attention”, “working memory”, “long-term memory”, and “perception”.

Higher-level cognitive functions include “speech & language”, “decision-making” and “executive control”.

These cognitive functions are processed and regulated by neural circuits that interact at the interfaces of the cerebral cortex, thalamus,

and basal ganglia. Solari and Stoner (2011) have collated, integrated, and visualized the accumulated connectivity data obtained from many published studies of primate cerebral cortex on neuroanatomical circuitry of cognition and integrated these fragmented studies into a single framework [50].

Based on this integrative model, the authors hypothesized that cognitive function follows a pattern of information flow among seven neural circuits. The seven circuits described are consolidated long-term declarative memory, short-term declarative memory, working memory/information processing, behavioral memory selection, behavioral memory output, cognitive control, and cortical information flow regulation (Figure 15).

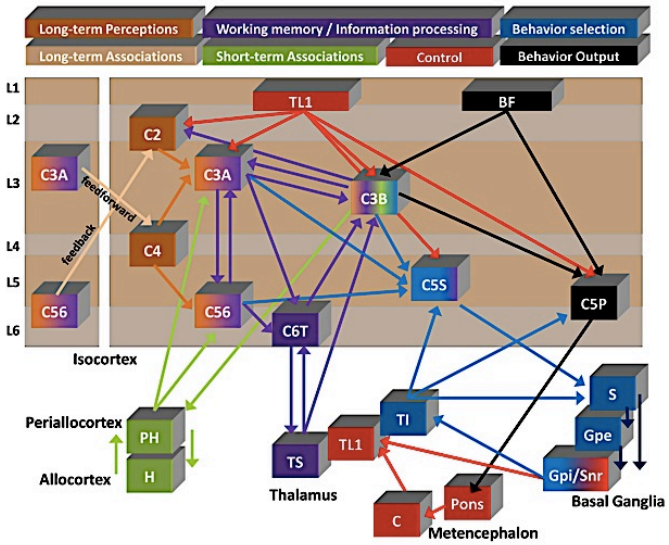


Figure 15. Summary diagram of proposed flow of cognitive information. Generally information flows from left to right through the color coded circuits. Long-term memory is split into perceptions and associations. Cortical neuron x (Cx), Parahippocampal gyrus (PH), Hippocampus (H), Specific thalamus (Ts), Layer 1 projecting thalamus (TL1), Intralaminar thalamus (Ti), Cerebellum (C), Striatum (S), External segment globus pallidus (Gpe), Internal segment globus pallidus (Gpi), Substantia nigra par reticulata (Snr), Basal forebrain (BF). (Taken from [http://www.frontiersin.org/files/cognitiveconsilience/ index.html#]; last accessed 01.11.2013)

To map cognitive functions to specific brain regions, functional brain imaging or electroencephalography can be utilized aiming at localization of function [51]. This has led to construction and representation of topological brain architecture in form of connectivity maps or networks (Figure 16). The nodes of these networks are usually inferred during tasks that manipulate one or more of cognitive functions.

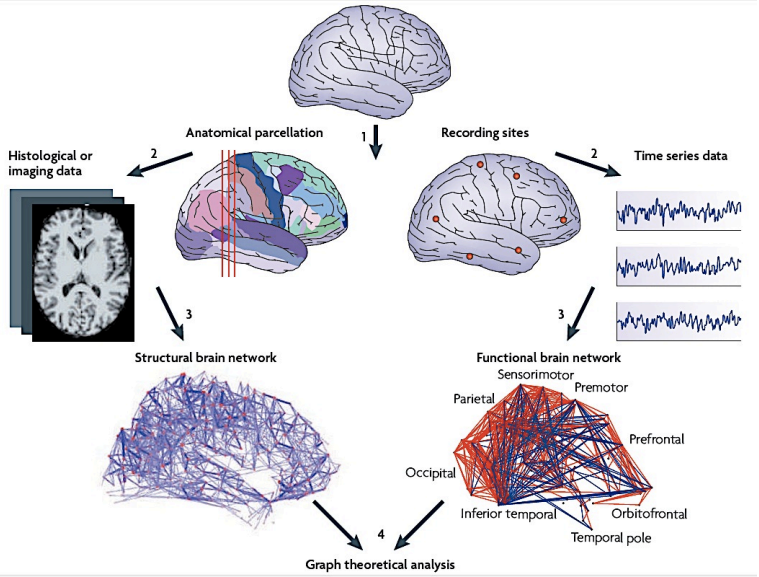


Figure 16. Structural and functional brain networks can be studied in four steps. 1: Define the network nodes as electroencephalography or multielectrode-array electrodes, or as anatomically defined regions of histological, MRI or diffusion tensor imaging data. 2: Estimate a continuous measure of association between nodes. This could be the spectral coherence between two magnetoencephalography sensors, or the connection probability between two regions of an individual diffusion tensor imaging data set, or the inter-regional correlations in cortical thickness or volume MRI measurements estimated in groups of subjects. 3: Generate an association matrix by compiling all pairwise associations between nodes to produce an undirected graph. 4: Calculate the network parameters of interest in this graphical model of a brain network and compare them to the equivalent parameters of a population of random networks. Taken from [51].

Alterations in the configuration of these connectivity networks reflect neural network function or dysfunction because cognitive and behavioral functions result from large-scale network interactions. For instance, fMRI-based connectivity networks distinguished between the Salience Network connectivity associated to behavioral variant frontotemporal dementia and the Default Mode Network connectivity associated to AD. While the former map represented atrophy in frontoinsular, cingulate, striatal, thalamic and brainstem nodes, the latter highlighted an atrophy specific to posterior hippocampus, medial cingulo-parieto-occipital regions and the dorsal raphe nucleus [52]. Therefore, functional and structural network mapping methods provide the opportunity to identify anatomically predictable patterns of neurodegeneration in the human brain.

2.3.2. Main hypotheses on the pathology of dementia

Since the early description of Alzheimer's disease made by Alois Alzheimer in 1906, and despite remarkable findings and insights into this disease, still the accurate mechanism of AD pathogenesis remains unclear. Several independent hypotheses have been proposed to address the pathogenesis in AD from different angles including apolipoprotein E (ApoE) genotyping [53],[54],[55], hyperphosphorylation of tau protein [56],[57], oxidative stress [58],[59], abnormal cell cycle [60],[61], inflammation [62],[63],[64], and Amyloid-beta metabolism [65],[66],[67]. In a recent review by Dong et al., these hypotheses have been systematically introduced and discussed [68]. Although all these hypotheses are important, none of them alone can explain the heterogeneity of pathological

abnormalities observed in AD. Here, I focus on three most popular hypotheses and briefly explain their proposed mechanisms.

2.3.2.1. Amyloid hypothesis

The hallmark of this hypothesis is the clinical observation of amyloid plaques and tangles in brains of deceased patients. According to this hypothesis, the imbalance between production and clearance of the amyloid-beta protein in the brain parenchyma leads to the deposition of toxic oligomers of this peptide, amyloid-beta aggregation, formation of tau filament aggregates and ultimately neural degeneration. Consequently, in both familial and sporadic AD production of total amyloid-beta or its 42 variant clearly increases but it is unclear whether the clearance of amyloid-beta in sporadic AD is disturbed [69]. Figure 17 summarizes downstream events that are shared between familial and sporadic AD as proposed by the amyloid hypothesis.

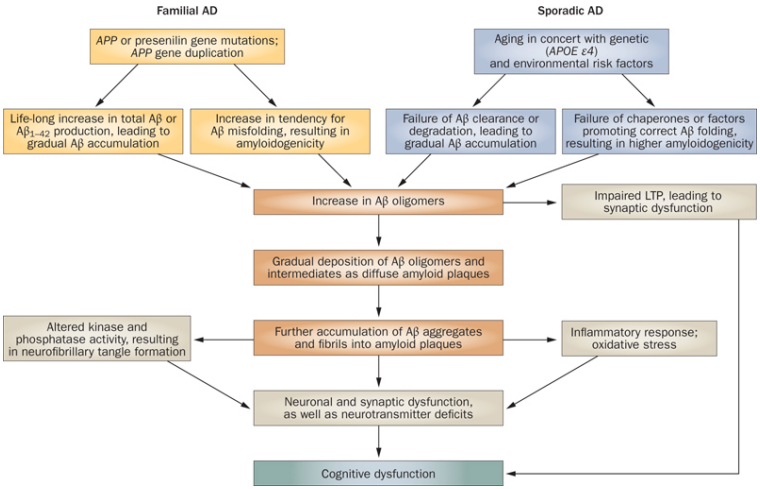


Figure 17. Diagrammatic representation of the amyloid cascade hypothesis in both sporadic and familial AD. An increase in production of either total Abeta or the amyloidogenic Abeta1–42 isoform is well established in familial AD, but only limited evidence exists for a specific disturbance in Abeta clearance in sporadic AD. In both familial and sporadic AD, soluble Abeta is thought to undergo a conformational change that renders it prone to aggregation into soluble oligomers and the larger insoluble fibrils found in plaques. Fibrillar Abeta deposited in plaques might be neurotoxic; however, synaptic loss and clinical progression of the disease mainly correlate with soluble Abeta levels. Soluble Aβ oligomers might inhibit LTP in the hippocampus and, hence, disrupt synaptic plasticity. Tau phosphorylation and subsequent neurofibrillary tangle formation, as well as inflammation and oxidative stress, are regarded as downstream events. Abbreviations: Aβ, amyloid-β; AD, Alzheimer disease; APOE, apolipoprotein E; APP, amyloid precursor protein; LTP, long-term potentiation. (Taken from [69])

The amyloid hypothesis explains the pathology mechanism as shown in Figure 18. The beta-amyloid precursor protein (APP) is cleaved by beta-site APP-cleaving enzyme (BACE) and the gamma-secretase complex. Consequently, the large domain of APP (CTF-beta) is secreted and binds to the complex of gamma-secretase, presinilin 1 or 2 (PS1, PS2). These all bind to each other and form a core complex required for gamma-secretase activity. The final products of gamma-secretase cleavage activity are amyloid-beta protein and the APP intracellular domain (AICD). The function of this peptide is not well understood [70].

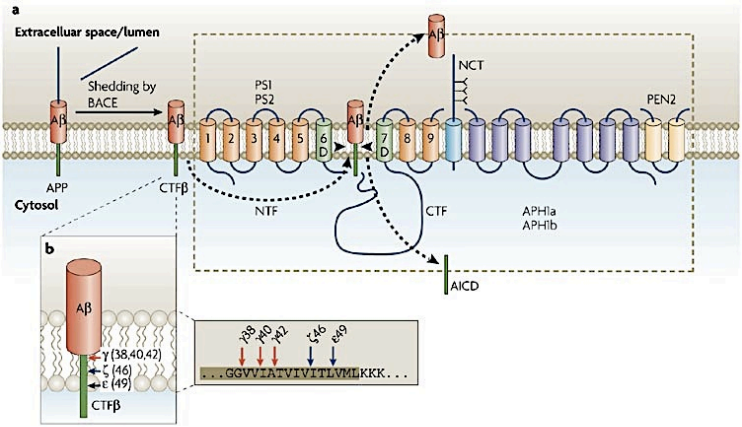


Figure 18. Cartoon representation of the molecular details underlying amyloidogenesis. a) Amyloidogenic processing of beta-amyloid precursor protein (APP) by beta-site APP-cleaving enzyme (BACE) and the gamma-secretase complex. b) Various proposed sites of intramembrane proteolysis by gamma-secretase. The amino-acid sequence around the cleavage sites of APP is shown (numbers refer to the sequence of Abeta; shaded amino acids are in the transmembrane domain). (Taken from [70])

Consequent failure of Alzheimer's drugs targeting amyloid-beta [71], on the one hand, and discovery of amyloid-independent neuronal injuries in demented patients [72], on the other hand, have recently led to calls for reconsideration of the amyloid cascade hypothesis [73],[74]. To this end, concerns have been raised that current anti-amyloid interventions in demented patients are too late [75]. New findings suggest that the pathological process in autosomal dominant AD begins more than 20 years prior to onset of dementia [76] and CSF levels of amyloid-beta increase up to 10 years before onset of late-onset AD [77]. Despite disappointing results from clinical trials of anti-amyloid compounds, Michael W. Weiner from Elan in his review article believes that "the research highlights of 2012 provide new support for the central role of amyloid in AD pathogenesis." [78]. Whether the amyloid-beta cascade is primary cause or secondary effect in pathogenesis of AD remains controversial.

2.3.2.2. Inflammation hypothesis

Neuroinflammation is a prominent feature of all neurodegenerative diseases. Frank-Cannon *et al.* (2009) provide a comprehensive review on neuroinflammation evidence in neurodegenerative diseases [79] but here I will focus on the role of inflammation in AD pathology.

First evidence of systematic inflammation involvement in the etiology of AD came from a meta-analysis study in 1996, which concluded - based on 17 epidemiological studies - that administration of non-streoidal anti-inflammatory drugs (NSAIDs) might reduce the

risk of AD [80]. Later, clinical observations showed that microglia accumulate surrounding amyloid plaques [81]. Thus, neuroinflammation is believed to be an early consequence of the amyloid beta deposition. Accordingly, aggregates of amyloid-beta activate microglia, partly through Toll-like receptors (TLRs), and these receptors activate transcription factors such as NFKB and AP-1, which in turn induce production of the reactive oxygen species (ROS) and drive the expression of inflammatory cytokines [82].

The rationale behind the inflammation hypothesis of AD is based upon the observation that both acute and chronic systemic inflammation were found to be clinically associated with accelerated cognitive decline in patients with AD [83]. A recent study shows that blockade of the IL-12/IL-23 pathway in a mouse model of AD led to a reduction in cerebral amyloid load and improvements in cognitive function [84]. Known molecular and cellular mediators of inflammation in AD pathology are summarized in Table 7.

Table 7. List of well-known mediators of inflammation in AD reported in biomedical literature

Cellular mediators	Microglia Astrocytes Oligodendrocytes Neurons
Molecular mediators	The complement system Cytokines and chemokines Toll-like receptors Cyclooxygenases and Arachidonic Acid Metabolites

Although Heneka and O'Banion suggest that inflammatory mediators may stimulate APP processing, and thus, anti-inflammatory treatment strategies should be considered for AD patients, still the precise molecular mechanism underlying this effect is not well understood [85].

2.3.2.3. Infection hypothesis

The first report on direct visualization and isolation of infectious agents from postmortem brain tissue of AD patients appeared in 1986. In this study, MacDonald identified Borrelia spirochetes in subculture of autopsy brain tissues from two patents with dementia [86]. Since then, there have been reports that hypothesized systemic infections may contribute to the pathogenesis of sporadic AD. Mawanda and Wallace (2013) now provide a comprehensive overview of infectious agents associated with AD and examine the latest status of the infectious AD etiology hypothesis [87]. These infectious agents include herpes simplex virus type 1 (HSV-1), Chlamydophila pneumonia, Borrelia burgdorferi, Helicobacter pylori, prions, and other infectious agents. I refer readers to the review paper mentioned above for more in-depth information.

Regarding the mechanism of action of infectious agents associated to AD, two possibilities have been proposed: either infectious agents directly infect the brain or indirectly promote the pathogenesis of AD through various effects of systemic infections [88]. Following infection, in addition to the inflammatory response, there is a systemic response (i.e. acute-phase response). Launch of this response is dependent on the synthesis of pro-inflammatory

molecules such as IL-1 beta, IL-6 and TNF. In a series of studies with transgenic mice, it was shown that the amyloid-beta protein has many of the characteristics of a prion, providing evidence for the existence of Abeta prions [89]. The hypothesis that AD might be a prion-type disease with possible capability of being spread by infectious particles was first raised by Morales and coworkers [90]. They injected brain extracts of AD patients into the brains of normal mice and were able to induce amyloid-beta deposition as well as the typical AD symptoms in control animals. At present, research on the impact of systemic infection on chronic neurodegeneration is ongoing and aims at better understanding of the communication routes between the peripheral immune system and the brain [91].

2.3.3. Clinical heterogeneity of AD

Heterogeneity in presentation and progression of AD has long been demonstrated (e.g. [92],[93] and [94].

There have been three hypotheses to explain such heterogeneity in AD: “subtype hypothesis” (distinct clinical variants), “phase hypothesis” (variation in the stage of the disease), and “compensation hypothesis” (variation in the origin and progression of the disease). Subtypes of AD have been classified based on the relative density of neurofibrillary tangles into “typical AD”, “hippocampal sparing AD”, and “limbic-predominant AD”. However, it has been shown that these subtypes are different than neurofibrillary tangle-predominant dementia, which might be more than a mere variant of AD [95]. Moreover, left parietal atrophy is associated with language symptoms in younger age of onset and

faster cognitive decline, which is indicative of a 'language variant' of AD, whereas right parietal atrophy is associated with visuospatial symptoms, suggestive of a 'visuospatial variant' [96]. The high degree of heterogeneity that is observed amongst AD patients can be also be attributed to the varying rates of disease progression. In a study of 80 AD cases, neuropathological heterogeneity was shown to be more consistent with the phase hypothesis [97]. Komarova and Thalhauser (2011) have quantified such temporal variations by mapping the variations onto the GDS/FAST staging system, and found that there is a large heterogeneity in the duration of stages in AD.

Understanding the neurobiological basis of this heterogeneity requires strategies that link cognitive and behavioral variations to the regional damage of brain cells. Imaging biomarkers have emerged as useful means for capturing variations in the distribution of the pathologic changes. They can also help clinicians distinguish AD from other dementia syndromes.

2.3.4. Treatment options for AD

Despite many years of research and gain of substantial insights into the pathology of AD, no disease-modifying drugs yet exist for AD. Early treatment studies focused on the inhibition of cholinesterases (AChEs) and its positive, but unspecific, effects on memory function. Tacrine, introduced in 1992, was the first such inhibitor, which was followed by three other inhibitors with better safety profiles, namely donezepil, rivastagmine and galantamine. Based on their demonstrated ability to stop cognitive decline for 6 to 12

months, these drugs were used to treat mild to moderate AD patients. The second wave of AD treatment began with memantine in 2003, which is an NMDA receptor antagonist and is the first drug approved for the treatment of moderate to severe AD. Interestingly, investigational drugs target almost all amyloid processing steps, as illustrated in Figure 19.

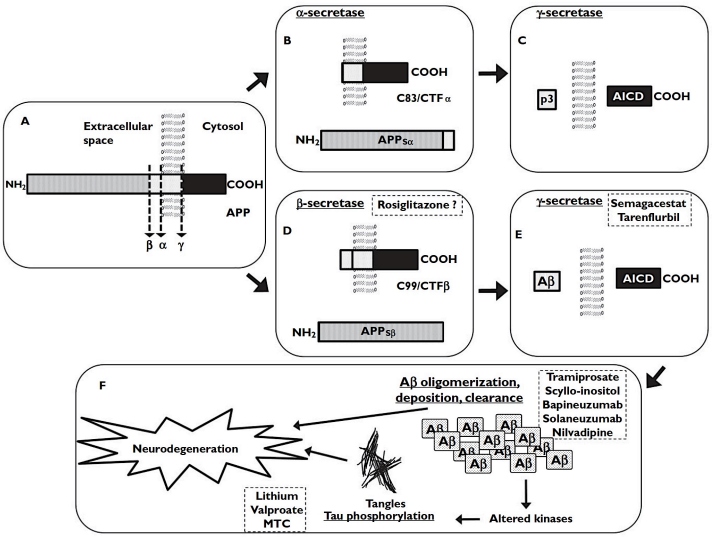


Figure 19. Main steps of sequential cleavage of amyloid precursor protein (APP), leading to generation of beta amyloid and/or other products with existing therapeutic interventions mapped to each stage. A: the cleavage sites for alpha-, beta- and gamma-secretase. B: ectodomain shedding of APP by alpha-secretase. C: Subsequent intramembrane proteolytic cleavage of C83/CTFα by gamma-secretase. D: ectodomain shedding of APP by beta-secretase. E: Subsequent intramembrane proteolytic cleavage of C99/CTFβ by gamma-secretase. F: Abeta

oligomerization and deposition leading to neurodegeneration, both directly and through tau hyperphosphorylation. In dashed boxes, disease modifying drugs that interfere with each particular step. (Taken from [98])

However, these treatments neither stop the neurodegeneration nor reverse the progression of AD. They have only symptomatic effects for a short term; e.g. they may improve over the baseline or delay to decline cognition but neurodegeneration inevitably continues its downward trajectory after 6 to 9 months. Besides, about 30-40% of AD patients do not respond to approved AChE inhibitors at all [99]. Accordingly, in the absence of disease modifying drugs that are able to slow or stop the disease, the current challenge of pharmacology is to develop the new generation of drugs that prevent AD. In this paradigm, it is expected that disease-modifying medications reduce the slope of decline or at least stabilize a patient's condition [100]. There are currently over 70 experimental compounds in preclinical and clinical trial phases, targeting mostly amyloid cascade or downstream responses such as inflammation and apoptosis [101].

CHAPTER 3. State-of-the-art Computational Approaches to Disease Modeling

3.1. Biological models and their types

In the philosophy of science, a "model" is an idealized structure that we use to represent the world, via relations between the model and real world target systems. The model is used by scientists to gain understanding of a complex real-world system via an understanding of simpler, hypothetical system that resembles it in relevant respects [102]. In biology, "a model is description of a system" and a system is "any collection of inter-related objects" that act as elemental units of the system on which observations can be made [103]. Since biological systems in the real world experiments are considered as complex, models of biological systems have been classified in the literature under the mathematical models. According to Barillot *et al.* (2012), two forms of mathematical models for biological systems exist: "data-driven models" versus "knowledge-based models" [104]. In data-driven models, the structure of model is inferred from the biological data whereas in knowledge-based models prior knowledge is compiled and incorporated into a model by human expertise. Accordingly, models can be categorized based on their types:

(a) quantitative or qualitative;

(b) statistical or mechanistic;

(c) static or dynamic;

(d) discrete or continuous;

(e) deterministic or stochastic.

These models are usually represented in the form of different formalisms as follows:

-Differential equation models

-Power law models

-Boolean models

-Petri-net models

-Rule-based models

-Flux balance analysis models

3.2. Modeling biological systems

Emergence of advanced technologies that enable researchers to obtain comprehensive snapshots of biological systems at real time and with single molecule resolution has led to the generation of unprecedented amount of data. However, the one-dimensional view, while useful, will fail to deliver models that are generally predictive of complex system behavior. *In silico* modeling and simulation of complex biological systems provides an efficient way to organize and integrate multiple data types and ultimately, to generate multi-dimensional views. Thus, systems biology can be called as "integrative biology" with the aim of predicting biological outcomes given the interaction of underlying components and elements [105].

Systems biology has emerged as a young discipline, which aims to investigate the structure and dynamics of biological systems (holistic approach) rather than the characteristics of the isolated parts (reductionist approach) [106]. Modeling complex biological systems

requires a good understanding of their features, amongst which hierarchy, complexity, heterogeneity, context-dependency, and emergent property are the most important features [107]. In fact, these features arise from the enormous interactions amongst molecular constituents of cells. As mentioned above, the challenge of systems biology is to understand the structure and the dynamics of intra- and inter-cellular web of interactions that contribute to the function of a living cell. The sum of these interactions forms a 'network of networks', which can be modeled and structurally analyzed. Barabasi and Oltvai (2004) were first to adopt the idea of network modeling from physics and introduce it to biology as 'network biology' [108]. This notion can be traced back to the application of graph theory to representing biological networks by Watts and Strogatz in 1998 who characterized the 'small world phenomenon' and other network properties [109]. More recent features and properties of complex biological networks can be found in the book entitled "Network science: complexity in nature and technology" [110].

3.2.1. Bottom-up systems biology

The bottom-up approach starts from constitutive parts and mechanistic details, and uses accurate formulating of each process to provide an interconnected view of biological interactions. The aim of this approach is to come up with an integrated map of pathway models representing the entire system. However, limitations of these approaches can be summarized as follows: i) complete dependency on experimental studies and data availability; ii) uncontrollable

enlargement of models by the incorporation of more data and processes; iii) vulnerable to experimental design and tools for model analysis.

Since the bottom-up approach follows the first principles in biology, it is not surprising that it requires extensive quantitative data, which is not always available particularly when a model of complex disease such as AD is going to be built. An example of bottom-up approach is modeling of enzyme kinetics, which requires determination of kinetic parameters by in-vitro measurements. The Silicon Cell model takes advantage of the bottom-up approach to enable online in-silico experimentation, for example, about the effect of over-expression of an enzyme in the pathway on the productivity of the interaction of interest [111].

Metabolic network models are often built using bottom-up approaches and typically include stoichiometry data. Computational stoichiometric models have been used to perform metabolic flux analysis as well as flux balance analysis including prediction of cellular phenotypes using extreme currents and extreme pathways [112].

3.2.2. Top-down systems biology

The top-down approach originates from large experimental data sets or information that provide structure to the model and tries to unravel underlying biological mechanism by narrowing the analysis down to the bottom. In this approach, the flow of information occurs from 'omics' level to pathway level. The top-down approach starts with data analysis and data integration to establish correlations among

molecules, followed by prediction and hypothesis generation. These hypotheses can be further tested in the lab and thereby, an iterative cycle is formed (Figure 20). The top-down approach uses various types of models at different phases of the characterization of the system; for example, in the component discovery phase, knowledge-based qualitative static modeling can be applied and in the later phases Bayesian models may be used to decipher the regulatory dependencies.

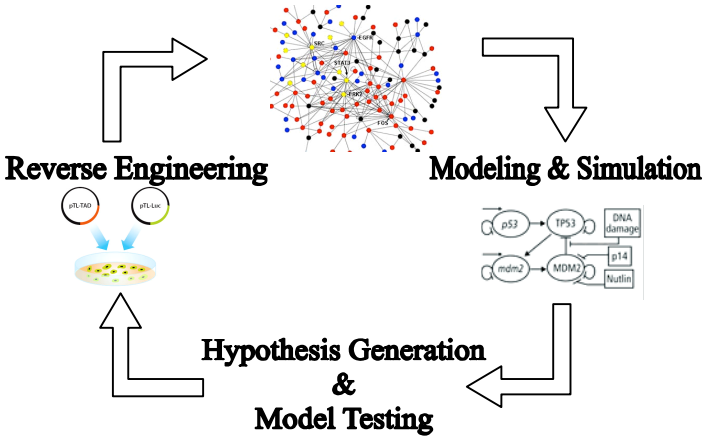


Figure 20. The cycle of modeling, hypothesis generation, and experimentation in systems biology. Systems biology is an integrated process, which begins with pattern observation in experimental data, continues with modeling/simulation of the observed pattern and leads to hypothesis generation. Generated hypotheses can be further validated by experiments.

Limitations of this approach include difficulty of dissecting many entangled mechanisms, correlation-based rather than cause-and-

effect-based, and lack of supporting knowledge for interactions. However, the advantage of this approach lies in its ability to identify network structures that underlie the system behavior through 'reverse engineering' of the system data. Such an ability allows for isolation of 'mechanistic structures' that explain the mode of action [113].

The bottom-up approach has been already applied to amyloid-beta accumulation in AD with the aim of elucidating mechanism of neurotoxicity [114]. This approach has led to the identification of cartoon-like pathways and mechanisms influenced by amyloid-beta including pathways in neurons, astrocytes and microglia that are implicated in beta-amyloid toxicity. To date, all the endeavors in this direction have been limited to generation of molecular disease maps such as AlzPathway [115]. However, since bottom-up studies are often based on data derived from transgenic animal models, they do not provide an accurate representation of therapeutic effectiveness in humans [116]. Hence, an integrated top-down approach may provide clues to determine pathological pathways that are causally linked to the disease phenotypes, on one hand, and are central to disruption of neuronal networks underlying the memory and cognitive function, on the other hand.

3.3. Network modeling of disease processes

Since its introduction to the field of biomedicine, network modeling has been used to better understand the effects of molecular interactions on disease progression. Although network modeling of disease pathways has been called differently by different authors

such as “network systems biology” or “systems pathology” or “network medicine”, it essentially paves the way for transitioning from ‘descriptive’, ‘correlative’ associations between molecular signatures and clinical phenotypes to ‘mechanistic’, ‘causal’ associations. Molecular networks that represent molecular states of the perturbed biological system underlying disease (also known as ‘disease maps’) provide a suitable framework for transitioning from descriptive to mechanistic mode by linking genetic information to disease processes and clinical phenotypes.

Based on the nature of data, various types of human interactomes can be modeled (genetic networks, RNA networks, protein networks, regulatory networks, metabolic networks). There still exist important caveats when constructing such network models: interaction data remains incomplete, there is bias towards particular interactions, and the available data is noisy. However, current interactome models, in spite of their incompleteness or biasedness, open up new avenues of research towards an integrated understanding of the pathophysiology of diseases.

3.3.1. Properties of disease network models

Properties of disease network models show fundamental differences compared to other network models such a social or technological network models. These properties can be classified into three types of modularity (Figure 21).

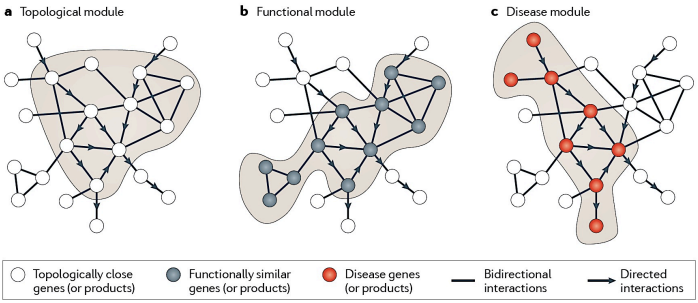


Figure 21. Comparative representation of three modularity concepts. a) Topological modules correspond to locally dense neighborhoods of the interactome, such that the nodes of the module show a higher tendency to interact with each other than with nodes outside the module. b) Functional modules correspond to network neighborhoods in which there is a statistically significant segregation of nodes of related function. c) A disease module represents a group of nodes whose perturbation (mutations, deletions, copy number variations or expression changes) can be linked to a particular disease phenotype, shown as red nodes. (Taken from [118])

While Zhu et al. (2007) provide a thorough review of the network models and their characteristics [117], I briefly touch the modularity feature of networks as follows.

3.3.1.1. Topological modularity

Molecular interactions are mathematically represented by graphs whose topological characteristics have been subject of extensive research to find shared design principles. Initially it was thought that the independence of scale or scale-freeness is indicative of networks that evolve gradually based on preferential attachment of nodes [108]

and accordingly, attempts were dedicated to matching topologies of molecular interaction networks such as gene co-expression, protein-protein interaction networks and metabolic networks to the topologies observed in other real-world networks (e.g. social networks). However, it was later figured out that properties of random network models are not consistent with properties of molecular networks in terms of statistical and biological meaning [119][120]. In parallel to these top-down analyses using degree distribution metrics, a bottom-up approach was utilized to find motifs and modules to describe the functional significance of biological networks but these descriptors are criticized for having no characteristic behavior when considered as a dynamic system [121].

In the context of disease representation, topological metrics attempt to answer the question whether disease genes/proteins are distributed randomly in the interactome or whether there are correlations between their locations and their network topological features. One special topological feature is the emergence of high-degree nodes or "hubs", which govern the regulatory processes in biological networks [122]. Based on biological role and dynamic behavior, hubs have been classified into "party hubs" and "date hubs", which function inside modules and between modules, respectively [123]. Dense connections between nodes with the same functionality form the so-called 'clusters' and are observed in highly modular networks. These clusters provide the basis for introduction of the concept of functional modules [124].

Although the current trend in topological network analysis is slowly moving from global to local analysis to couple topological properties

to biological knowledge, the main caveat associated to topological features is that they may not reflect the real biological modularity as a network model is assumed to be always incomplete. Given this and other above-mentioned reasons, for the rest of this thesis work, I am cautious to rely on topological metrics in my analysis and will only focus on the functional modularity concept.

3.3.1.2. Functional modularity

A functional module is defined as "an aggregation of nodes of similar or related function in the same network neighborhood." [125]. Functional modules carry out cellular functionality and are hierarchically organized [126]. The aim of functional module analysis techniques is to identify functional pathways that regulate sub-processes.

Early functional analysis algorithms were focused on mapping out the gene ontology (GO) concepts to nodes and testing the enrichment of annotation terms iteratively. This approach, which suffers from the large linear output of annotations terms was further developed into the widely used gene set enrichment analysis (GSEA) approach. The main advantage of GSEA is that it is a 'knowledge-based' approach, in the sense that it focuses on manually curated gene sets that share common biological functions [127]. Modular enrichment analysis approach takes advantage of traditional ontological analysis but incorporates extra network discovery algorithms [128]. Since disease-specific networks generated and presented in this thesis work are large, the GSEA algorithm will be applied to perform functional analysis.

3.3.1.3. Disease modularity

Disease modules are disease components or subnetworks whose dysregulation leads to disease phenotypes. In the context of network models of diseases, disease-related components are inter-related. The key difference of the disease modules compared to the topological and/or functional modules- despite their possible overlap- is that they are specific to the disease they represent, meaning that each disease can have its unique module. Since disease modules include disease-modifying genes, they are likely to explain multiple disease phenotypes in both sporadic and hereditary diseases [118].

Disease module-based methods start with building a disease-specific cellular interactome, identifying subnetworks with most of the disease-associated genes, and validating the functional role of the module in the pathology of the disease using functional annotations or expression patterns [129][130]. Often the limitation of these methods is the low coverage of disease-specific maps in the vicinity of the known disease components. However, such methods help us dig into disease mechanisms and develop rationales and hypothesis that could be used to guide experimental research towards more objective outputs.

3.3.2. Applications of disease-specific network models

Analysis of disease network models can provide a rationale approach to guiding systems interpretation in the context of prior physiological or pharmacological knowledge along with incorporation of both specific (target) and non-specific (off-target) effects. Besides

provisioning an integrative view, disease models can be used for biomarker and drug discovery purposes, as I briefly highlight below.

3.3.2.1. Disease mechanism discovery

Despite the availability of huge body of information and knowledge on human diseases, the molecular mechanisms underlying complex human diseases remain largely unexplained. The challenge of linking clinical outcomes to their underlying molecular events has been long of interest to the scientific community and to the pharmaceutical industry as well. This is because it will help to obtain better understanding of the disease mechanism at molecular level. Thus, disease-specific models may be used to support translation of molecular findings into the development of new therapeutic strategies.

Traditionally, disease-specific maps are built based on correlation network models. Such models are routinely built using high-dimensional data such as protein-protein interactions or gene expression data by establishing pairwise relations (i.e. edges) for all variables (i.e. nodes) but they confound direct and indirect associations and do not distinguish between cause and effect. Alzheimer's disease map [115] and Parkinson's disease map [131] are good examples of disease-specific networks in the area of NDDs. In contrast, causal network models aim at representing response variables and covariates and, thus, the directionality of associations between cause and effect. When relationships between variables represent conditional dependencies (e.g. given disease symptoms, compute the probabilities of the presence of various diseases), the

model is a Bayesian network, which requires information on prior distribution. In the absence of such information, Biological Expression Language (BEL) offers an alternative. BEL is the state-of-the-art causal network modeling language that integrates literature-derived 'cause and effect' relationships into a data-driven platform and produces casual network models (The Biological Expression Language [http://www.openbel.org]). The models developed by this means represent a high-resolution and comprehensive view not only on the core established pathways but also on peripheral events that lead to clinical readouts seen in patients. A prime example is development of mechanistic disease models for identification of patients with ulcerative colitis who could be potential responders to targeted anti-TNF therapy with Infliximab [132]. Based on the prior knowledge in the literature, a causal network model was constructed that described mechanistic knowledge underlying ulcerative colitis in the form of the "cause-relationship-effect" pattern. Next, gene expression profiles of responders and non-responders were incorporated into the causal model and a mechanistic strength value was calculated on the gene expression network activity signature of TNF signaling for each patient in the population. The model demonstrated that non-responders have different TNF signature compared to responders, which was due to sustained TNF-like downstream signaling in non-responders after treatment with Infliximab, controlled by alternative upstream controllers (Figure 22).

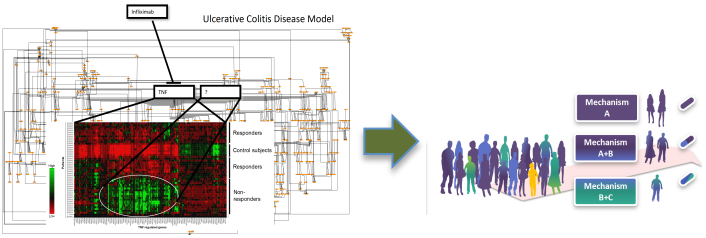


Figure 22. Stratification of patient population based on disease mechanism. A BEL model was successfully built and applied to a cohort of patients with ulcerative colitis disease – based on TNF signature - to stratify responders from non-responders to Infliximab.

It is noteworthy that a computer-readable disease model should be readily amenable to computational reasoning for disease mechanism discovery and furthermore, in the absence of models that represent normal cellular processes in healthy state, any attempt to derive mechanistic interpretations is inconclusive. Realizing this necessity, Westra et al. (2011) constructed a computable BEL model to capture normal, non-diseased biology of cell proliferation in lung cells [133]. "Differential network analysis" (also known as "differential network biology") methods have been devised and used for cross-species, cross-condition network analysis since the 1990s [134]. However, despite the fact that biological systems are highly dynamic and change depending on environment, tissue type, disease state or developmental stage, to date, almost all network models have represented the cellular interactions under single static condition (mostly under disease condition). Although conventional approaches

try to map out dynamic changes in gene expression or metabolic fluxes on static network models, they are unable to capture changes in wiring of the network model. To tackle this limitation, several studies have begun to use disease-specific models for modeling changes across disease states. A prominent example is the analysis of dynamic modules in the human protein interaction network that was shown to be useful for the prediction of breast cancer outcomes [135].

Network perturbation amplitude (NPA) measurement is an up-to-date and elegant approach that enables modeling of mechanistic network re-wiring and quantitative measurement of network perturbation in response to external stimuli [136]. The NPA scoring method integrates high-throughput experimental data (e.g. gene expressions) into curated literature-derived knowledge in the form of BEL models, provides an scoring function for quantification of causally affected biological processes, and quantifies the changes of disease state in comparison to control (non-perturbed) state. Although the applicability of this scoring function to other areas has not yet been demonstrated independently since its introduction in 2012, it promises to compute the amplitudes of treatment-induced perturbations in biological networks.

3.3.2.2. Drug-target identification

Low productivity of drug discovery pipelines in recent years has been largely attributed to insufficient efficacy of failed drugs [137]. Failure of drugs in phase II and III clinical trials in fact reflects the poor understanding of the mode of action of such drugs at the

molecular level. Network-based knowledge of disease mechanism can be used to identify disease modules and prioritize drug targets. This approach is essentially a top-down approach, which aims at providing a network-based view of drug action (polypharmacology) while keeping the focus on node-based target prediction. Such methods have been successfully applied mostly to cancer and metabolic diseases [138],[139] and are currently under investigation for NDDs [140],[141].

Earlier approaches considered topological characteristics of disease networks for target identification such as hub nodes, bridging nodes, etc. However, knockout studies revealed an unprecedented 'robustness' in biological networks that leads to masking phenotypes due to functional compensation of neighboring pathways. For this reason, drug target prediction algorithms shifted towards the identification of perturbations rather than single genes [142]. The latest network-based target identification methods rely on targeting a particular group of proteins, being complexes, modules or pathways [143].

Given the above background, two network-based strategies for target identification has been proposed [144]:

1) "The central hit strategy", which aims at targeting hubs, central nodes, or master-regulators of infectious disease networks [145] as well as cancer (e.g. in ovarian cancer [146]), and the goal is to selectively hit the network integrity of the infectious agent or the malignant cell; and
2) "The network influence strategy", which relies on the systems-level knowledge of both healthy and diseased states

to identify target candidates in polygenic complex diseases so that the ultimate object is to reverse the disease network malfunctioning to its normal functioning.

Methods of the network influence strategy are much less developed than the single hit strategy as they deal with the dynamics of the disease network and it is a difficult task to overcome the robustness of networks. Beside nodes, edges of disease networks can be targeted to enforce the rewiring of the disease modules. The so-called "edgetic drugs" open up new opportunities to modify or modulate interactions of multifunctional proteins with a larger selectivity [147]. Current efforts in relation to targeting edges have been mostly concentrated on protein-protein interactions (PPIs) and it is expected that these efforts extend to signaling as well as gene interaction networks.

3.3.2.3. Biomarker discovery

Disease-specific network models play a substantial role in both data-driven and knowledge-driven methods of biomarker discovery. In the data-driven approach, similarities between data points (typically gene expression data) are computed in order to identify network clusters, which can suggest mechanisms of regulation. In contrast, knowledge-based approaches take advantage of the expert knowledge embedded in context-specific curated PPIs, functionally annotated pathways, and biomedical publications. Given the unique advantages of both data- and knowledge-driven approaches, therefore, a combined approach to network-based biomarker discovery is preferred [148].

Azuaje et al. (2012) suggest two integrative biomarker discovery strategies: a) linking network structure properties to clinical phenotypes, and b) identification of subnetworks with differential molecular response [149]. For instance, integration of gene expression profiles obtained from primary breast tumors into a protein interaction network model led to identification of sub-network biomarkers that represent metastatic tumor progression [150]. The philosophy behind using subnetwork-based versus node-based biomarker discovery is that molecular changes of individual nodes represent different outcomes in different network contexts of different patients. On the basis of the same philosophy, differential network analysis methods are being devised to identify subnetwork biomarkers (biomarker modules) by integrating stage-specific microarray data into disease-specific and control-specific PPI networks; for example, Liu et al. (2012) present a novel approach to predicting rewiring of network interactions in three stages of gastric cancer [151].

The advantage of network-based biomarker discovery methods is that they have greatly facilitated the paradigm shift from "signature-based biomarkers" to "multi-panel (multiplex) biomarkers". By this means, gene expression information is linked to proteomics and pathophysiology of disease so that many different types of biomolecules (i.e. a panel of biomarkers) are being associated to pathological processes [152]. For instance, Dudley and Butte (2009) created two biomarker networks by linking genomic profiles from human diseases to detectable biomarker protein in blood plasma and in urine, respectively [153]. They found that in both networks, over

80% of the putative biomarkers are linked to multiple disease conditions. Thus, the vast combinatorial space of candidate biomarkers could be computationally investigated using a top-down approach, compared to the conventional bottom-up approach.

3.3.2.4. Classification of diseases

The contemporary classification of human diseases is based on observational correlations between anatomical, symptomatic and epidemiological criteria, which does not take the etiological mechanism into account and thus, lacks sufficient sensitivity and specificity. This has led to the problem of misdiagnosis due to the overlap of symptoms. Consider Parkinson's disease with dementia and Alzheimer's disease dementia. Since classification of these two neurodegenerative diseases is merely based on their clinical signs, early diagnosis of Parkinson's disease (PD) is hampered. The extent of clinical overlap between these two conditions is so much greater than chance that even some authors propose that AD and PD belong to a spectrum of neurodegenerative disorders with common disease mechanism but triggered by different etiological factors [154]. However, at the molecular level, these two conditions can be distinguished based on diagnosis between synucleinopathy from tauopathy.

Another shortcoming of the current disease classification system is ignorance of inter-connected nature of many diseases, including comorbidities. Mechanistic disease models can be used to reveal inter-dependencies between the disease and risk mechanisms (Figure

23). Such interdependencies are reflected in comorbidities that co-occur with the disease and increase the disease risk.

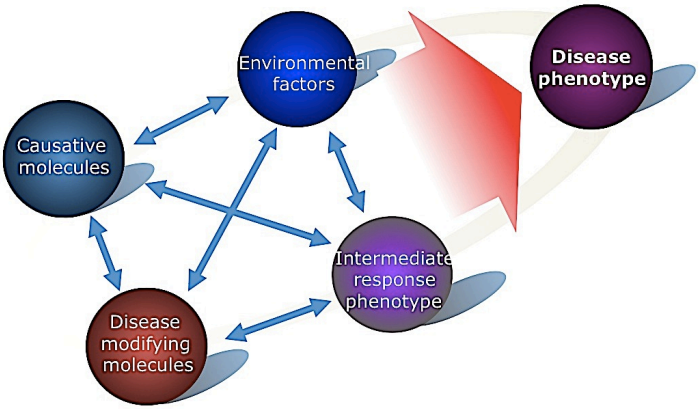


Figure 23. Deterministic factors involved in shaping the disease phenotype. Complex interactions between causative molecules, disease modifying molecules and environmental factors lead to an intermediate phenotype response (e.g. as comorbidity), which contributes to the manifestation of disease endophenotype.

Thus, integrative disease modeling provides a systems-based network framework for incorporation of both conventional reductionism and non-reductionism approaches and allows for re-definition of the current nosology [155]. In 2011, National Academy of Sciences committee recommended NIH to develop a new taxonomy of diseases based on their underlying molecular and environmental causes rather than on physical signs and symptoms [156]. The idea is to create a so-called Knowledge Network information system that integrates molecular data, medical histories,

and health outcomes of individual patients. In Europe, the Innovative Medicines Initiative (IMI) - Europe's largest public-private partnership aiming to improve the drug development process – launched a call that addresses the topic of developing an aetiology-based taxonomy for human diseases. The aim of this call is to propose new etiological mechanism-based taxonomy in immunoinflammatory diseases and neurodegeneration with focus on AD and PD by developing knowledge frameworks that represent clinically relevant mechanistic knowledge of the pathophysiology underlying neurodegeneration.

3.3.2.5. Drug side-effect prediction

Mapping drugs onto the target space of disease networks provides a system-wide view of the target landscape, which can be used for the investigation of polypharmacology as well as side effects of off-targets. Network pharmacology has been proposed to aid selection of those compounds that maximize modulation of disease networks while minimizing side effects [157]. In a polypharmacological analysis, Keiser et al. (2009) computed chemical similarities among FDA-approved drugs and predicted new off-targets that explained drug's adverse reactions [158]. In fact, both drug-target networks and drug-disease networks can be used for the prediction of side effects [159],[160]. A landmark paper by Campilus et al. (2008) showed that computing phenotypic side-effect similarities could predict targets that are shared between drugs and can be used for drug repositioning [161]. Mitzutani et al. (2012) analyzed correlations between drug target-binding profiles and side effect profiles in the context of a

drug-protein interaction network and found that correlated sets were significantly enriched with proteins that are involved in the same biological pathways [162]. In summary, network prediction methods may help to decipher novel side-effects when analyzing the disease-specific mechanistic networks for identification of potential drug targets.

CHAPTER 4. Setting an *In-Silico* Strategy for Optimizing Early Drug Target Identification and Validation

4.1. Target selection in drug discovery

Target identification in drug development is a crucial step for follow-up medicinal chemistry studies. A target in the context of drug discovery can be defined as disease-associated proteins that are functionally involved in the pathology of interest. Distinctions are typically made as to whether a target is 'novel', 'established', or 'validated'. Briefly, novel targets are proposed targets with speculative involvement in the disease process and no clear indication of its clinical benefit whereas established targets are those with a good scientific support on functionality in both normal and disease states but unknown clinical benefit. In contrast, validated targets have shown a clear clinical benefit with a well-understood mechanism of action.

Specific biological hypotheses based on which targets are selected possess varying degrees of confidence, depending on the origin of those hypotheses. Given that targets have been historically identified on the basis of genetic studies or biological observations, Sams-Dodd from Boehringer Ingelheim Pharma (2005) divides targets into three classes: "physiological targets" that characterize physiological effects at the level of whole organism in animal models; "genetic

targets" that represent genetic mutations; and "mechanistic targets" that represent receptors, enzymes, or other biological molecules and are linked to molecular mechanism [163]. Accordingly, genetic targets are specific to those diseases that arise from a genetic mutation or the increased disease risk by a single gene. Moreover, the gene or its product must be the main modulator of the disease at the time of intervention. These two conditions for genetic targets imply that multifactorial complex diseases that develop over time can not be treated by such an approach. Instead, mechanistic targets go beyond the 'single gene, single disease' paradigm by engaging environmental factors in addition to causative biological components, and therefore, can be applied to multifactorial progressive diseases. Since mechanistic targets affect multiple molecular mechanisms, their validation is a complex task and depends on the availability of predictive models.

In general, there are three complementary approaches to target identification: biochemical methods, genetic interaction methods, and computational prediction methods [164]. Amongst these three methods, computational target identification methods have the advantage of the least bias compared to other methods because they rely on a combination of experimental data generated by others. A high-priority task for computational target identification methods is to address the issue of clinical efficacy; i.e. in the absence of clinical data, a model is required to integrate both the experimental data and expert knowledge in the context of the disease of interest so that ultimately the clinical efficacy of a target can be predicted *in silico* in the form of a set of relevant hypotheses [165]. A success story in this

regard is the instrumental role of computational data integration and model analysis in identification of Aurora kinase A as the key target of dimethylfasudil in acute megakaryoblastic leukemia. In this case, integrating transcriptomic and proteomic data led to generation of testable hypothesis, which identified relevant target of dimethylfasudil [166].

4.2. Model-driven approach to target identification and validation for neurodegenerative diseases

The pressing need for treatment of neurodegenerative diseases has not yet met due to complexity of disease mechanism and multiple unknown genes and pathways. The majority of existing treatments were discovered empirically and, for this reason, the mode of action of most of these drugs was unknown when they were introduced to the market. Consequently, the conventional target-based approach has not been much successful in the area of NDDs as was for other disease indications [167]. This fact is supported with the current and subsequent failures of Alzheimer's drugs, as discussed in Chapter 1. Therefore, a better and deeper understanding of the neurodegeneration mechanism is needed, if the target-based approach continues to play its role in the CNS drug discovery and development. In response to this need, pharmaceutical industry has already taken steps towards applying systems biology approaches to deciphering disease mechanism and identifying therapeutic targets for CNS disorders, including NDDs, as discussed by Mei et al. from GlaxoSmithKline in their seminal review [168].

Normally, a research process starts with an exploration of the problem domain by collecting relevant data, information, and previous knowledge, which are often hidden in scientific publications. Accordingly, referring to the scientific literature is usually the first step towards drug target selection and validation process [169] because it provides a valid and proper framework for drug target identification purposes. When merged with network-based disease models, the information extracted from text enhances the confidence about druggability of the candidate target(s). Moreover, it would be possible to generate informative profiles for each candidate target using information extracted from the text; i.e. literature-based annotation of target nodes on the network model of disease provides enormous insight about drug candidate efficacy and toxicity. Such profiles will be of high value for ranking or prioritizing target candidates.

4.2.1. The overall strategy

To address the motivation and mission of this research, and to improve our current understanding about pathophysiological mechanisms underlying AD, an "integrative disease modeling" approach is proposed, which takes advantage of the complementary nature of data-driven and knowledge-driven methods, combines them under a single framework, and produces knowledge-based, yet mechanistic disease models. The models generated by this approach could represent correlation or cause and effect, depending on the type of associations between pairs of variables in the network model. The general strategy is depicted in Figure 24.

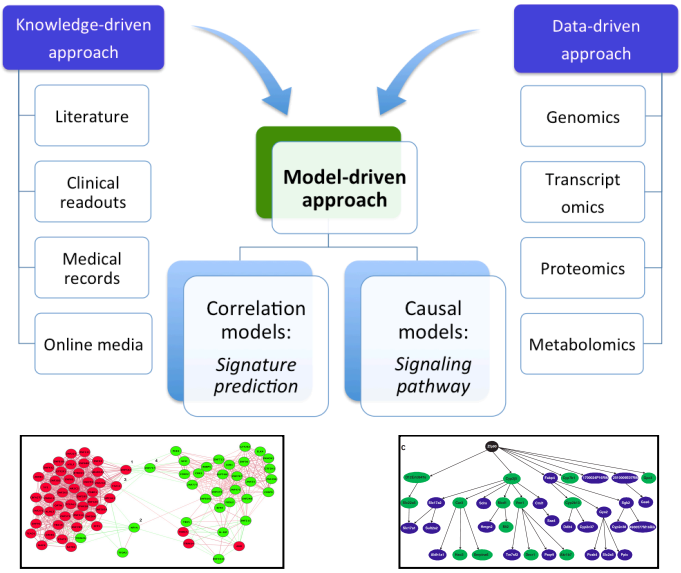


Figure 24. Model-driven approach to integrating biological data. Proposed model-driven approach combines biomedical knowledge and data into a single disease model, which could represent correlations (left) or cause and effect (right).

4.2.2. The methodology

In the following, I discuss the workflow that was implemented according to the above strategy. It is a biphasic approach, meaning that knowledge-driven and data-driven methods are applied in two independent phases (Figure 25). Network models built from protein interaction data are used as model backbones or integration templates for addition of literature-derived information as well as overlay of experimental data. As discussed in Chapter 3, correlation-based networks including PPI networks have both advantages and disadvantages to be used as the model backbone: while PPI networks

lack the directionality on edges and suffer from incompleteness, their biggest advantage over the other types of networks is that they enable us to link data-derived information to literature-derived knowledge. Moreover, PPI networks offer the flexibility required for mapping other biological data types such as gene expression.

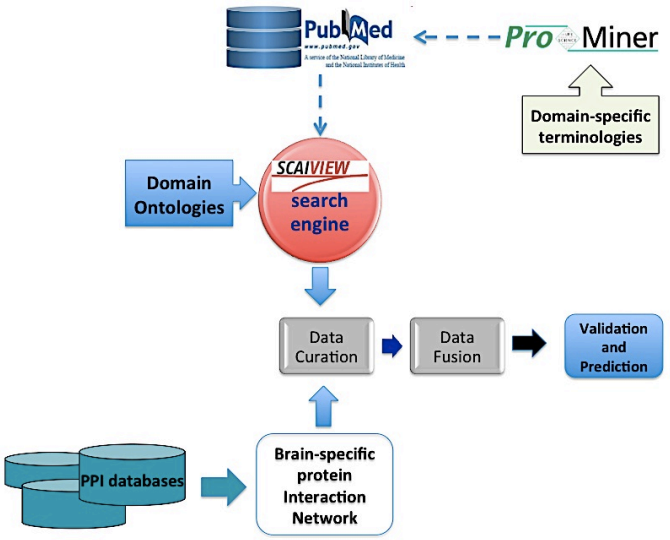


Figure 25. Diagrammatic sketch of the workflow applied to the modeling approach. Domain-specific terminologies were plugged into the ProMiner software and the results of named entity recognition on PubMed abstracts were deployed in the SCAIView environment, appearing as highlighted text. The harvested knowledge from ontology-supported queries was manually integrated with protein interaction data using a brain-specific network model to isolate brain region-specific subnetworks for further analysis.

4.2.2.1. Knowledge-driven approach

Text mining has been defined as "the computational discovery of new, previously unknown information, by automatically extracting information from different written sources" [170]. Text mining technologies have been broadly used for the identification of disease-related knowledge on potential targets, biomarkers and the disease mechanism.

Generally, text-mining methods consist of two steps [171]: information retrieval (IR) and information extraction (IE). IR is performed with the aid of search engines via two approaches: rule-based (knowledge-based) and statistical (machine-learning). The former uses pattern recognition to find a meaningful biological pattern such as '<brain> and <pathway>' whereas the latter uses classifiers to classify abstracts or sentences. IE relies on named entity recognition (NER), which enables detection and extraction of biological entities such as genes, SNPs, proteins, drugs, or other entities. Numerous tools have been developed on the basis of NER techniques: iHop [172] annotates a subset of PubMed sentences containing at least two proteins in conjunction with interaction-specific keywords. AliBaba [173] aggregates results of a PubMed query and visualizes them as a graph. EbiMed [174] retrieves co-occurring entities and ranks them by frequency. UKPMC [175] annotates and highlights entities in PubMed Central abstracts by using Whatizit [176]. GoPubMed [177] recognizes entities such as genes, gene ontology terms and MeSH terms. GeneView [178] is rather a recent NER-based tool, which recognizes a broader set of

entity types than genes but not gene ontology terms, provides search facilities using unique database identifiers and also finds relationships between proteins in text. ProMiner [179] is one of the systems for gene normalization, which performed very well in BioCreAtIvE I and II assessments, reaching an F-score of 0.8 for the recognition and normalization of human genes and proteins. The ProMiner system has been designed for the semi-automated generation of dictionaries and the recognition of spelling variants, as well as the disambiguation of acronyms and common word synonyms. ProMiner uses extensive dictionaries for various biological entities, including genes/proteins, SNPs, chemical names, disease terms, and any other relevant vocabulary.

IE methods aim at extraction of relevant entities from the retrieved documents in a tabulated form. The simplest IE approach identifies co-occurrence of entities in the text whereas a more complex IE method attempts to extract relations between entities such as protein-protein interactions, drug-target interactions or biological pathways. Several IE tools have been developed, which analyze search results and present summarized knowledge of semantics: MedEvi [180] provides concept variables of major biological entities (e.g. gene) to be used in semantic queries and prioritizes search results based on keywords that occur in original queries. EBIMED [181] extracts proteins, GO annotations, drugs and species from retrieved documents and identifies relationships between extracted concepts based on co-occurrence analysis. CiteXplore [182] provides integrated access to both literature and biological data and also contains abstract records from patent applications from the Europe

Patent office and from the Chinese Academy of Sciences. MEDIE [183] provides semantic search in the format of triples (subject, verb, object) and returns abstract sentences that match the queried semantic relations. PubNet [184] parses the XML output of standard PubMed queries and creates different kinds of networks. Nodes can be representatives of article, author or some database IDs and edges are constructed based on shared authors, MeSH terms or location. Ontology-based information extraction systems have been emerged as a subfield of IE, which make the use of ontologies and their explicit conceptualization to semantically guide the extraction of classes [185].

The literature retrieval system used in this thesis is composed of two software components: the named entity recognition tool ProMiner and the knowledge discovery framework SCAIView [186]. The retrieval terminologies are incorporated into ProMiner, followed by ProMiner's annotation of all Medline abstracts using the dictionaries. The resulting entity annotations are stored in an Apache Lucene(TM) - based search index, together with the documents and their meta-information. SCAIView incorporates the Lucene-based index and allows for searches that include exact matches, wildcard options, and Boolean operations. The document visualization within SCAIView provides highlighting for all the entity classes with tooltips containing available linkouts, descriptions, and depictions. The annotations are organized in the form of hierarchies (semantic trees) and can be navigated by the user via selecting full classes (e.g. ‘genes’ or ‘diseases’), selecting certain subclasses (e.g. ‘Nervous Systems Diseases’), or singular dictionary entries (e.g. ‘Alzheimer’s

Disease'). For Boolean operations, either single entity class or complete tree classes could be selected. The system returns the results according to statistical ranking of found entities based on Kullback-Leibler divergence (relative entropy), meaning that the more relevant entities (e.g. gene names) appear in the top of the ranking list.

Ontologies go beyond the mere providers of a common terminology for a knowledge domain by taking into consideration the semantic relationships among the concept classes. Such relationships help to model the knowledge domain in a well-organized hierarchy of concepts, which add to the value of ontologies when it comes to the real world applications. In contrast to philosophical ontologies, computational ontologies aim to represent a scientific knowledge domain in a machine-readable taxonomy-based manner with the maximum proximity to the natural language expressions in the real world. Computational ontologies form the basis for knowledge-driven drug discovery approaches. To this end, I have integrated dictionary-based text mining technologies and ontological search capabilities to retrieve and extract relevant information for disease modeling and drug target identification (Figure 26).

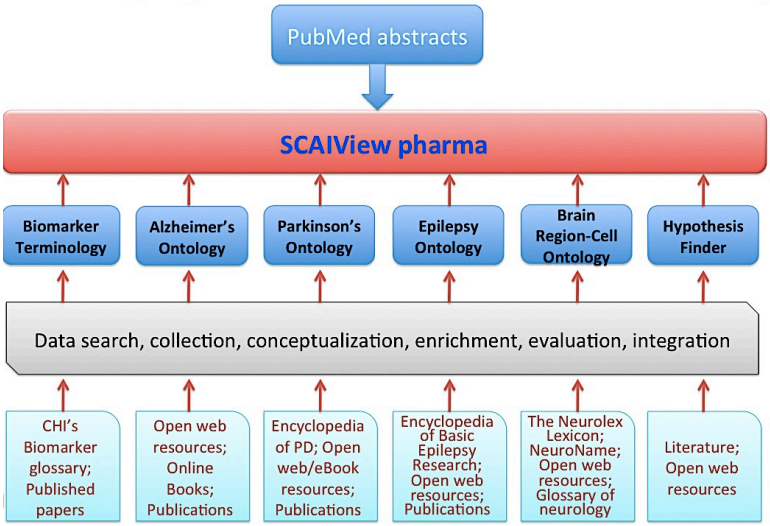


Figure 26. Schematic workflow of ontology-driven literature search. Various resources were used to build ontologies and their controlled vocabularies, which ultimately were plugged into the SCAIView advanced search engine to allow for semantic search.

The advantage of this approach is that a combination of domain ontologies and their dictionaries can be used in SCAIView to narrow down the search in the vast space of the knowledge cube. Therefore, the user is able to perform targeted, advanced search and export entities together with their annotation (Figure 27). The Entity View displays a table with the aggregated list of entities found in the documents ranked by relative entropy, as explained before. In the

Entity table, the following columns are shown by default: Entity, Relative Entropy, Reference Documents Count, Documents Count, and Link-outs to databases. However, this table can be expanded by adding more annotation columns to include additional information such as links to KEGG or Reactome pathways as well as Gene Ontology, InterPro family and domain information, ATC (Anatomical Therapeutical Chemical Classification) code, etc. The full table as well as selected entities can be exported the text file (CSV format). It is possible to export the list of all the extracted entities (e.g. gene names) along with their corresponding PubMed identifier so that there is a means for tracing back the reference from which the entity has been extracted. Exporting co-citations as well as PPIs out of the selected entities are other potential IE functionalities of SCAIView.

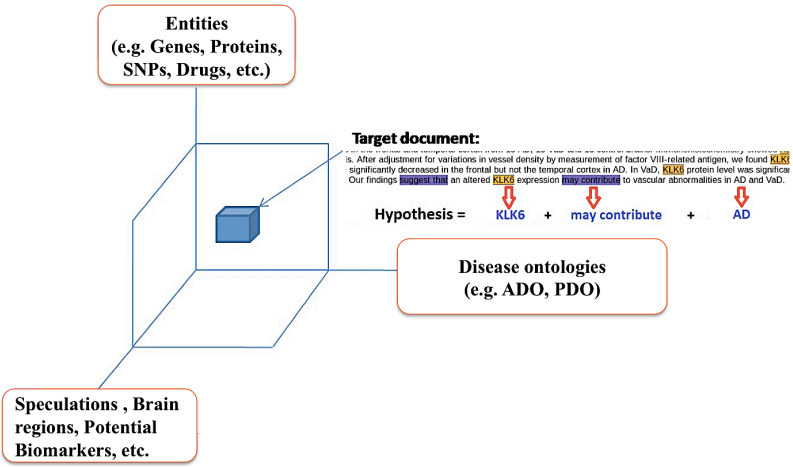


Figure 27. Ontology-driven IR and IE using SCAIView facilitates targeted search.

The published knowledge in the literature contains complex biomedical information that can be represented as a multi-dimensional cube. Controlled vocabularies derived from ontologies are used in combination to explore the knowledge cube in multiple dimensions and perform targeted search for informative documents.

4.2.2.2. Data-driven approach

With the availability of novel high-throughput methods, large amounts of data about the interactions among proteins are available in various databases. Aggregation of PPIs under a single interactome is necessary because there is a small overlap amongst PPI databases in terms of their data content [187],[188]. For this reason, several tools have been devised to capture, collect, and unify PPI data from existing PPI databases. For instance, BIANA (Biologic Interactions and Network Analysis) integrates multiple sources of biological information, including proteins and their relationships, and attempts to manage the biological information as a network where proteins are nodes and interactions are edges [189]. The PPI networks generated by means of such automatic methods, however, contain noise (e.g. incomplete parsing of data points, erroneous normalization of node names, etc.) and must undergo expert curation process. During this process, additional annotations on nodes and edges of the network model are performed using the knowledge-driven approaches. For instance, nodes that are already targeted by approved drugs can be highlighted or edges representing interactions in specific tissue or cell type can be filtered. The visualization, statistical analysis, and filtering of PPI networks is carried out in the environment of the Cytoscape software [190].

4.2.2.3. Model building workflow

Both the information retrieved and extracted from the knowledge-driven approach and the data aggregated from data-driven approach are subjected to a dedicated workflow with curation, quality control and model assembly steps (Figure 28). For instance, the protein interaction data are curated for any possible parsing error during the process of automatic data collection, aggregation and ID conversion. In this work at the reformatting step, all the protein IDs were converted into the official gene symbols as defined by HUGO Gene Nomenclature Committee (HGNC). By this means, mapping of other data layers such as gene expression levels or expressed biomarkers for the purpose of model analysis becomes feasible. Similarly, the textual information retrieved and extracted from publications undergo a manual curation and quality control process to avoid false positives and to ensure the true relevance of extracted information to the corresponding query. This step is particularly important as the current state-of-the-art IR and IE tools do not produce noise-free results and thus, intervention of human expert is still required.

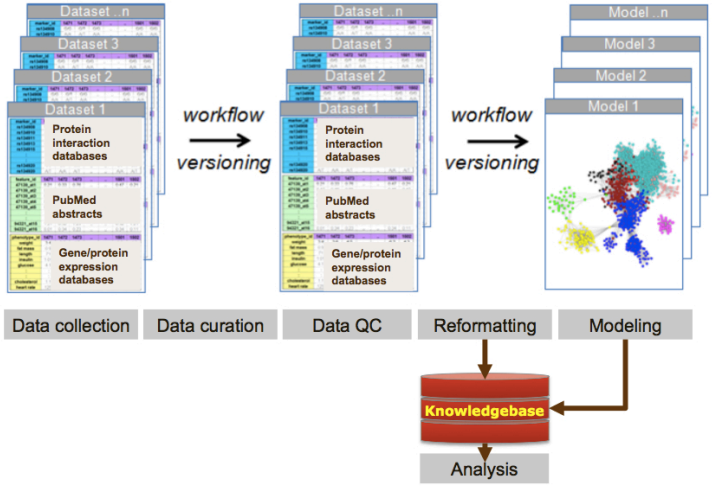


Figure 28. Schematic workflow of model building. This workflow was used to construct context-specific models using curated, quality controlled data types.

Once models are built, they are deposited in a dedicated knowledgebase together with proper documentation and meta-information for the sake of future provenance. During analysis stage, functional validation of generated models is performed through pathway analysis. As explained in Chapter 3, GSEA methods are often used to determine enriched functional modules in the network model. Recovery of manually curated pathways from databases such as KEGG or BioCarta is another alternative. In the next chapters, I introduce a novel network validation method entitled "biomarker-guided pathway analysis", which takes advantage of the accumulated

knowledge about potential molecular biomarkers of disease and makes use of this observational information to guide the pathway analysis to the core molecular events already experimentally recorded for the pathology of the disease.

Once the most significant pathway modules are identified in the context of tissue- or stage-specific disease network, these modules can be prioritized to identify candidate pathway modules as described in Chapter 10 (Figure 29).

Figure 29. Workflow for validation, integration, and analysis of subnetwork models. Context-specific integrated models were subjected to validation by a variety of strategies such as biomarker guided analysis and pathway recovery test to establish the relevance and specificity of the model. Validated model is then used for identification of important disease processes supported by evidence attributes.

CHAPTER 5. Multi-scale Knowledge Representation using Ontologies

Biological and medical ontologies are formal representations of biomedical knowledge. They have been proven very useful for the communication of biomedical information through controlled vocabularies, definitions and proper metadata annotation [191]. Numerous examples have demonstrated their value for data mining and knowledge discovery approaches. Ontologies have been used for automated reasoning [192], for large-scale annotation of entire genomes [193], for data mining in microarray data [194] and for semantic and ontological search in unstructured information sources such as scientific text [195],[196].

The biological domain has developed a large portfolio of widely accepted and widely used ontologies. However, relevant knowledge in the pharmaceutical sector has not yet been addressed by the public scientific community. Some proprietary efforts to organize the knowledge relevant for pharmaceutical industry exist. To our state of knowledge, the BioWisdom pharma ontology [197] is the only ontological resource representing a substantial part of the pharma world. This proprietary ontology comprises substantial evidence (extracted from literature) and incorporates a broad spectrum of public sources.

Building context-specific disease models requires domain-specific knowledge to be incorporated into the model. Since disease models

may span from molecules to cells, tissues, organs and the entire physiology of human, there is a need to build up ontologies that map to these biological scales [198]. In the absence of such semantic resources for NDDs, I developed several ontologies to support the knowledge-driven approach for modeling mechanistic disease networks, with focus on Alzheimer's disease (Figure 30).

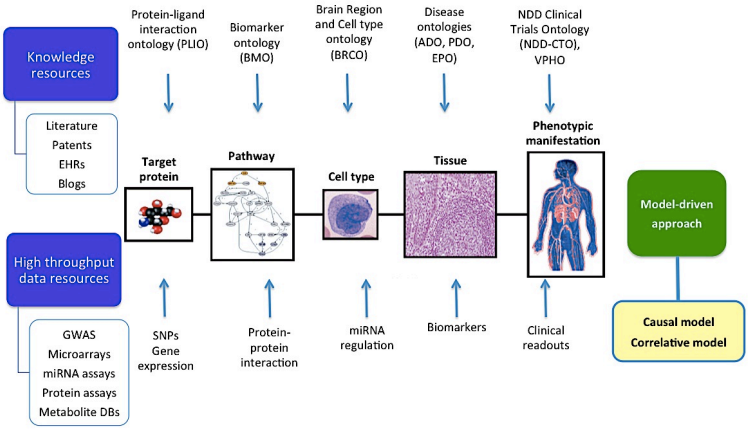


Figure 30. Mapping ontology development efforts to various biological scales. Different ontologies and controlled vocabularies shown in this diagram were developed to support IE and IR process for building context-specific systems models.

5.1. Ontology life cycle

The ontology-building life cycle is a common standard practice and is composed of the following five steps, as illustrated in Figure 31 [199].

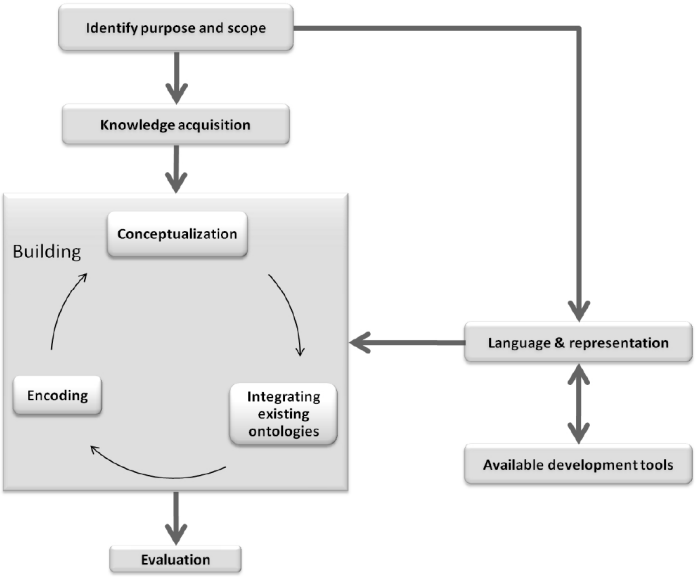


Figure 31. The life cycle of ontology-building consisting of five substantial steps. A detailed description of the ontology-building life cycle is outlined below.

1) Defining ontology purpose and scope

The first step in an ontology construction process is the definition of an ontology purpose and an ontology scope. In the ideal case, before the ontology building process begins, it should be clear why the ontology should be built. For defining an ontology purpose, the

ontology builder should identify motivating scenarios and competency questions. Also the motivating scenarios can help to identify the range of possible users. The set of competency questions should help to define ontology scope.

2) Knowledge acquisition

Knowledge acquisition consists of knowledge gathering from web, articles, books and other knowledge resources. An analysis of the collected knowledge should be performed with respect to the ontology scope.

3) Conceptualization

When enough knowledge is aggregated - the conceptualization process takes place, which consists of defining concepts and relationships between them. During this step, the superclass-subclass hierarchy of the ontology concepts is established. Parts of other existing ontologies can be reused and integrated into the ontology under development.

4) Encoding

During encoding step, the concepts are coded into the OWL (Open Web Ontology Language) structure format using specialized software tools, such as Protégé (Protégé Project [http://protege.stanford.edu/]).

5) Ontology evaluation

The last step is the evaluation of the ontology. Gangemi et al. (2006) have defined three measurements for quality assessment of ontologies:

I. *Structural measurement*, which measures topological and formal semantics of the ontology;

II. *Functional measurement*, which tries to assess the coverage and applicability of the ontology to the knowledge domain of interest;

III. *Usability profile measurement*, which aims to measure the extent of efficient accessibility of users to knowledge levels.

Structural evaluation is performed by calculating features such as depth, breadth, and other topological features of the ontology. To evaluate the functional quality of the ontology in terms of measuring the boundaries of the knowledge domain it captures, precision, recall, and F-score values are calculated [200].

Precision is the number of true positives (TP) divided by the sum of TP and false positives (FP).

Recall is the number of TP divided by the number of results that should have been returned (FP + false negatives (FN)).

The F-score = 2 x (precision x recall) / (precision + recall).

These values are derived from the longest string match found between automatically annotated words using ProMiner and the human-curated gold standard annotation for each abstract in the selected corpus.

5.2. PLIO: Protein-Ligand Interaction Ontology

The PLIO ontology was built in the course of a Master thesis as a groundbreaking effort to practice ontology-based IR and IE and calibrate the text-mining machinery and the IR system. This work has been published with equal contribution of the author of this thesis [201] and it is mentioned here for the sake of scientific

completeness of the work done for identification of novel targets at the molecular level.

5.2.1. Purpose and scope

PLIO is representing knowledge about the interaction of proteins and ligands (incl. drugs) and has a different scope and conceptual resolution than the Molecular Interaction ontology [202]. An important feature of PLIO is that it links directly from an ontology framework describing protein-ligand interactions to the mathematical formulas relevant for the computation of some of the entities represented in the ontology. To our knowledge, this is the first example for an ontology, which directly links from a knowledge representation to the mathematical building blocks that describe the leaves of the ontology in mathematical terms. It is noteworthy that we have adopted the top-level formal ontology structure during the construction of PLIO, i.e. the principle criteria of the top-level ontologies using the Basic Formal Ontology (BFO) upper level concepts was followed [203]. However, our attempt was concentrated on keeping the concept definitions close to expressions in natural language. Thus, the hierarchical structure of the ontology can serve as a robust navigation tree for terminology integration and text-mining applications.

5.2.2. Structural characterization:

Assessment of the quality of PLIO was based on both, structural and functional criteria. Table 8 summarizes the structural features of the ontology.

Table 8. Structural parameters of the PLIO ontology and their corresponding values

	Diameter	Depth	No. of concepts	No. of leaves
Classes	371	13	371	271
Properties	13	0	12	12

5.2.3. Functional characterization:

In order to assess the functional quality of the PLIO, three competency questions were sketched. Answering the competency questions requires sufficient ontological coverage to capture the concepts of the domain. Table 9 represents the competency questions that can be answered by the PLIO.

Table 9. List of competency questions, the range of concepts that each question covers, and corresponding relations defined in the ontology

Competency question 1	**Concepts**	**Relationships**
Find features which reflect protein-ligand interactions from the viewpoint of biophysics, chemoinformatics, molecular modeling, and experiments.	a) Interaction descriptor b) Interaction type c) Thermodynamics of protein-ligand interactions d) Interaction detection e) Interaction simulation	-Interaction *"has_a"* Interaction descriptor -Interaction *"has_a"* Interaction type -Interaction *"described by"* Thermodynamics of protein-ligand interactions -Interaction *"part_of"* Interaction Detection -Interaction *"part_of"* Interaction simulation
Competency question 2	**Concepts**	**Relationships**
Find features which reflect a certain kind of ligand	a) Intrinsic activity b) Binding activity	-Ligand "*has_a*" activity -Ligand "*has_a*" ligand

activity against its biological target.	c) Biological activity d) Conformer	binding site -Ligand "*binds to*" ligand binding site -Ligand "*is_a*" conformer -Ligand "*has_a*" conformation -Activity "*is_a*" intrinsic activity -Activity "*is_a*" binding activity -Activity "*is_a*" biological activity -Activity "*has_a*" activity landscape -Activity "*has_a*" structure-activity relationship -Activity "*is_a*" biotransformation ability -Activity "*is_a*" reaction ability
Competency question 3	**Concepts**	**Relationships**
Find features which reflect the properties of protein and ligand binding site.	a) Ligand binding site b) Ligand binding site properties	-Ligand Binding site "*has_a*" binding site properties -Binding Site properties "*is_a*" physicochemical properties -Binding Site properties "*is_a*" geometrical properties

After enrichment analysis of the training set, 81 concepts were enriched with synonyms and 25 new concepts were added to the protein-ligand interaction ontology. The terminology behind PLIO supports 1321 synonyms (on average 3.5 synonyms per concept). Evaluation of the terminology showed a satisfactory performance on an independent test corpus of 100 Medline abstracts (Table 10).

Table 10. Results of the ontology evaluation using NLP-based approach

	Precision	Recall	F-score
Independent test set of 100 abstracts	0.94	0.72	0.8154

5.2.4. Usability profile

PLIO provides users with 1036 entity annotation axioms for all instances and classes. The coverage of relevant information in the ontology has been increased by adding 75 formula annotations and several software hyperlinks. Through integration of PLIO in SCAIView, we could make the ontology easily navigable as a tree and – at the same time – visualize the markup of PLIO concepts tagged in PubMed abstracts.

5.3. BMT: Biomarker Terminology

During the past years, high-throughput technologies have been extensively employed for the study of molecular mechanisms underlying different diseases and this has led to the discovery and development of a large number of molecular biomarkers. Several definitions for biomarkers have been proposed amongst which the one by the US National Institutes of Health defines a biomarker as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention”.

Biomarkers have been utilized throughout various stages of drug discovery and development. For example, biomarkers play an important role in drug target discovery and validation (e.g. as quantitative readouts for candidate drugs), in the monitoring of toxicity mechanisms (e.g. quantitation of indicators for unwanted side-effects), and non-invasive imaging of diseased organs. In the process of drug development, biomarkers are considered to be pivotal to informed decision-making as they are used to drive critical go/no-go decision in the early stages of drug development.

For target identification purposes, potential biomarkers reported in the literature indicate measurable molecular activities under the disease condition, which provide more reliable evidence in-vivo at the protein level compared to inconsistent gene expression results. This has important implications in the study of NDDs because gene expression data are often obtained from the post-mortem brain, suffer from heterogeneity of cell-types and tissues mixed in the brain samples, and are not even disease stage-specific. Hence, later in the course of this thesis, I introduce the novel concept of 'biomarker guided pathway analysis' by which I guide the disease model analysis for target identification in AD to core pathological processes underlying disease mechanism. A first step to finding supportive evidence for clinically important potential biomarkers is to search the accumulated data and knowledge generated from basic research. For efficient exploration of the suspected large amount of biomarker information contained in the biomedical literature, semantic search and information retrieval systems are of utmost importance. Hence, development of a biomarker terminology as a semantic resource for

systematic harvesting of AD-related potential biomarkers was performed and published [204].

5.3.1. Purpose and scope

Surprisingly, there exists no ontology for knowledge representation of the biomarker domain. The purpose of BMT was to capture all the key concept classes in the domain of biomarkers under a single classification schema to be used for text-mining applications. The scope of BMT covers six classes based on the conventional classification of biomarker types as well as the distribution analysis of the potential biomarker information in the literature.

1. *Clinical Management*: annotates all terms indicating clinical investigations in patients, which includes the initial mentioning, the clinical study, and finally the treatment
2. *Diagnostics*: annotates all diagnostics that are used, which includes the initial disease stage, the molecular identification, and blood diagnostics
3. *Prognosis*: annotates all terms indicating any prognosis, outcome, or marker (e.g. clinical or biomarkers, adverse effects, resistance, response, disease progression or outcome)
4. *Evidence marker*: annotates all changes in gene and protein abundance, spanning from expression to mutation, SNP variations to phosphorylation status
5. *Antecedent*: annotates all risk factors mentioned for the relevant disease.

5.3.2. Structural characterization

The biomarker terminology contains 119 entity classes with 1890 different synonyms.

5.3.3. Functional characterization

The BMT was evaluated against a gold standard list of Alzheimer's disease genes. Genes were ranked based on frequency and evaluated against the Alzheimer's gold standard. For the different selections recall, precision, F-score and rank have been estimated for 10, 30, and 50 % recall. In addition the maximal recall has been estimated (Table 11).

Table 11. Performance evaluation for Alzheimer's disease. Genes were ranked based on frequency and evaluated against the Alzheimer's gold standard at different ranking positions.

Selection	Rank	Recall	Precision	F-score
Baseline: Genes / Proteins	60	0.10	0.92	0.17
	230	0.30	0.73	0.42
	469	0.50	0.61	0.55
Maximal F-score	728	0.67	0.52	0.59
Baseline + Clinical Management	61	0.10	0.90	0.17
	226	0.30	0.75	0.42
	465	0.50	0.61	0.55
Maximal F-score	682	0.65	0.54	0.59
Baseline + Evidence Marker	62	0.10	0.89	0.17
	225	0.30	0.75	0.42
	464	0.50	0.61	0.55
Maximal F-score	654	0.62	0.54	0.57

Baseline +Prognosis	63	0.10	0.87	0.17
	247	0.30	0.68	0.41
	541	0.50	0.52	0.51
Maximal F-score	740	0.61	0.47	0.53
Baseline + Diagnostics	64	0.10	0.86	0.17
	237	0.30	0.71	0.42
	494	0.50	0.57	0.53
Maximal F-score	520	0.52	0.57	0.54
Baseline + Statistics	64	0.10	0.86	0.17
	227	0.30	0.74	0.42
	678	0.50	0.42	0.45
Maximal F-score	377	0.41	0.62	0.49
Baseline + Clinical Management + Evidence Marker	60	0.10	0.92	0.17
	224	0.30	0.75	0.42
	451	0.50	0.63	0.56
Maximal F-score	555	0.57	0.59	0.58
Baseline + Clinical Management + Evidence Marker + Prognosis	61	0.10	0.90	0.17
	230	0.30	0.73	0.42
	568	0.50	0.50	0.50
Maximal F-score	479	0.47	0.56	0.51

5.3.4. Application scenarios

To demonstrate the applicability of the biomarker retrieval terminology, the performance of retrieval for biomarker-related abstracts from Medline was tested in the field of Alzheimer's disease. The system was able to successfully extract candidate biomarker genes/proteins relevant to the queried diseases (see table 12 below). Genes retrieved by the selection of Clinical Management and Evidence Marker classes and not mentioned in the AD gold standard were checked manually by this author using the SCAIView environment. Such genes/proteins might be valuable for

identification of novel biomarkers because they represent the yet-to-be identified biomarkers whereas those already matched with the gold standard are candidates that are better known as potential indicators of the disease. For this purpose, the abstracts must contain at least the information of the gene, which is altered (e.g. overexpression or mutation) in a particular state of AD and their therapeutic response state. Examples of such information with the corresponding PMIDs are given in Table 12.

Table 12. Examples of articles accepted to contain biomarker information. Example evidence for genes retrieved via SCAIView for Alzheimer's disease but not found in the corresponding gold standards.

PMID	Gene	Alteration	Textual Evidence
17387528	ACAD8 HMGCS2	SNP	In a European screening sample of 115 sporadic AD patients and 191 healthy control subjects, we analyzed single nucleotide polymorphisms in 28 cholesterol-related genes for association with AD. The genes HMGCS2, FDPS, RAFTLIN, ACAD8, NPC2, and ABCG1 were associated with AD at a significance level of P < or = 0.05 in this sample.
17531353	SLC17A7	Protein expression decrease	Loss of VGLUT1 and VGLUT2 in the prefrontal cortex is correlated with cognitive decline in Alzheimer disease…We quantified VGLUT1 and VGLUT2 in the prefrontal dorsolateral cortex (Brodmann area 9) of controls and AD patients using specific antiserums. A dramatic decrease in VGLUT1 and VGLUT2 was observed in AD using Western blot

19863188	HPX SERPINF1	Cerebrospinal fluid concentration	Five differentially-expressed proteins with potential roles in amyloid-beta metabolism and vascular and brain physiology [apolipoprotein A-1 (Apo A-1), cathepsin D (CatD), hemopexin (HPX), transthyretin (TTR), and two pigment epithelium-derived factor (PEDF) isoforms] were identified. Apo A-1, CatD and TTR were significantly reduced in the AD pool sample, while HPX and the PEDF isoforms were increased in AD CSF

For AD, out of 400 genes, 158 genes/proteins had at least one evidence in the literature as being a potential biological indicator of Alzheimer's disease and thus were considered as true positives (~ 40%). Evaluation of retrieved genes not existing in the gold standard for AD showed that almost half of these genes have probably the potential of being considered as candidate biomarkers. This indicates that automated text mining using biomarker terminology combinations increases recall for biomarker – specific information retrieval and makes it possible to systematically explore the biomarker space in an efficient way. As I explain in the next chapter, I have used the BMT utility to harvest and screen all potential biomarkers of AD reported in the literature.

5.4. BRCO: Brain Region and Cell Type Ontology

Modeling details of disease mechanism in NDDs is challenging due to the anatomical complexity of the human brain. To the state of our knowledge, preceding to this work, there was no semantic framework in the public domain for representing the heterogeneous

structure and anatomical granularity of the human brain spanning from the gross regions of the brain down to region-specific cell types. Building this semantic resource supported knowledge-driven model building with the aim of increasing model resolution and specificity at the molecular level while maintaining the link to the cell type and tissue level.

5.4.1. Purpose and scope

Currently there are two ontologies whose knowledge scope is close to the human brain anatomy, including Neuroscience Informatics Framework Standardized Ontology (NIFSTD) [205] and NeuroNames [206]. NIFSTD is a well organized and highly standardized ontology which covers a broad range of neuroscience domain, but for extensive knowledge retrieval related to a specific human brain region or cell type its usage is restricted by its limited terminology. On the other hand, although NeuroNames is considered as the largest source of brain region names, it does not cover the knowledge scope of human brain cell types. Moreover, both of these ontologies are not human-specific. In contrary, BRCO is a human-specific brain anatomical ontology, which enables users to traverse through the brain partonomy to cellular level with a coherent granularity.

5.4.2. Structural characterization

The structural features of BRCO reflecting topological and logical properties were measured by means of context-free metrics including depth (related to the cardinality of paths in a graph), breadth (related to the cardinality of paths), tangledness (No. of multi-hierarchical

nodes/ No. of all the nodes, a measurement of organizational fitness), and fan-outness (related to the dispersion of leaf node sets). These features are summarized in Table 13.

Table 13. Structural parameters of the BRCO ontology and their corresponding values

	No. of roots	No. of classes	No. of synonyms	Max. depth	Average width	Tangled-ness	Fanout-ness
BRCO	3	3238	9843	16	211.7	0.02	0.85

High value of Fan-outness factor (0.85) illustrates that BRCO has a very broad coverage. Tangledness factor of 0.02 shows presence of relatively very few multi-hierarchical nodes, which indicate that BRCO is easy to maintain.

5.4.3. Functional characterization

Applying ProMiner, evaluation of the terminology showed a satisfactory result on an independent test corpus (Table 14), although more iteration of ontology enrichment with false negative concepts will probably lead to improvement of the complete match.

Table 14. Evaluation of functional aspect of the BRCO ontology using text-mining approach. Complete match indicates capture of the complete concept term whereas partial match refers to partially matching concept terms due to their combinatorial representation with other words in text.

BRCO	Precision	Recall	F-score
Complete match	0.71	0.69	0.69
Partial match	0.81	0.80	0.80

5.4.4. Application scenarios

One particular scenario for the application of BRCO is generating brain region or cell-type specific gene signatures for NDDs based on information retrieved from both literature and microarray databases such as GEO. We used an expression profiling study in GEO database (GSE5281) characterizing gene expression levels between AD vs. normal aged human brains. This study considers 6 distinct brain regions with about 14 biological replicates per brain region. However, no detailed information on the outcome of the study can be found due to the poor description of the gene expression studies in the microarray databases. Hence, to further gain mechanistic insight into brain region- or tissue-specific biological processes for those dysregulated genes, BRCO was used to find lines of evidence from literature. Through the literature search, visual cortex (VCX) showed to be relatively spared from AD pathologies. Thus, we expect to see less expression differences between AD-affected VCX and healthy

ones, which are consistent with our observation that statistically significant genes identified by comparison of affected VCX and healthy ones showed significantly smaller logged fold-changes compared to other regions under investigation. The literature-derived enriched gene specific to other five regions were compared to their corresponding experimentally-derived regional enriched genes and common genes signatures were detected, as shown in Table 15.

Table 15. Comparison of literature and experimental approach towards regional enriched genes under AD. The low number of overlapping genes indicates the wealth of un-annotated data in microarray data repositories.

Brain regions	**No. of enriched genes derived from literature**	**No. of enriched genes derived from microarray data**	**No. of overlapping genes**
Entorhinal cortex	198	1932	32
Hippocampus	1020	3839	266
Middle temporal gyrus	38	5421	17
Posterior cingulate cortex	14	3457	5
Superior frontal gyrus	43	979	3

It is noteworthy that despite the great discrepancy observed between the literature-derived enriched genes and the genes derived from

expression profiling on the single gene level, we observed a better similarity at the level of the enriched biological processes (GO terms).

Another application scenario was to demonstrate that BRCO ontological search could be used to find confirmatory statements as well as experimental evidence supporting the microglia involvement in AD pathology. Microglia are often found near damaged tissue in Alzheimer's disease patients, but their role in the pathology of AD is not well known. Until very recently, the only hypothesis regarding the involvement of microglial cells in Alzheimer's disease (AD) pathogenesis is centered around the notion that activated microglia are neurotoxin-producing immune effector cells actively involved in causing the neurodegeneration [207]. Table 16 lists documents retrieved based on nested-ontology search with “microglia” concept from BRCO and “Alzheimer’s Disease” concept from the disease ontology in SCAIView. It could be demonstrated that not only established knowledge statements of molecular mechanism within microglia under AD were successfully retrieved (proof of concept), but also experimental evidence regarding subclasses of microglia could be retrieved by such approach.

Table 16. Comparative study of established knowledge statements and knowledge gained by BRCO for the role of microglial cells in AD. The extra knowledge gained with the help of BRCO complement the established knowledge for the role of microglia in all the given three examples.

Established knowledge recovered from literature	**Supporting knowledge gained from literature by BRCO**
Activated microglia have contradictory roles in the pathogenesis of AD, being either neuroprotective (by clearing harmful Aβ and repairing damaged tissues) or neurotoxic (by producing proinflammatory cytokines and reactive oxygen species). (PMID: 21763676) Microglial cells play an important role in mediating neuroinflammation in Alzheimer's disease (AD) by production of a series of proinflammatory mediators and clearance of Aβ peptides and senile plaques. (PMID: 21496499) Early microglial accumulation in Alzheimer's disease (AD) delays disease progression by promoting clearance of β-amyloid (Aβ) before formation of senile plaques. However, persistent Aβ	Our findings using both Iba1 and antiferritin immunostaining of microglial cells show that coincident with the appearance of tau pathology in DS subjects there is consistent presence of dystrophic microglial cells and conspicuous absence of activated microglia using both markers. (PMID : 21847625) These results suggest that MyD88 deficiency may reduce Aβ load by enhancing the phagocytic capability of microglia through fractalkine (the ligand of CX3CR1) signaling and by promoting apoE-mediated clearance of Aβ from the brain. (PMID : 21763676) In mouse models of AD, bone marrow-derived microglia (BMDM) have been shown to

accumulation despite increasing microglial numbers suggests that the ability of microglia to clear Aβ may decrease with age and progression of AD pathology. (PMID: 18701698)	delay or stop the progression of AD and preventing their recruitment exacerbates the pathology. (PMID: 21418002)

5.5. HuPSON: human Physiology Simulation Ontology

It is foreseen that modeling and simulations will provide a better understanding of the human's body functioning and its pathological processes and help to develop therapies and tools that can aid disease diagnosis, treatment and prevention. In order to support these actions, we developed, evaluated and published an initial version of the Human Physiology Simulation Ontology (HuPSON) [208] as a resource that supports meta-data annotation of simulation as well as fundamentals of modeling and links the systems biology work to the Virtual Physiological Human (VPH) [209] and the CellML initiatives [210]. This ontology has been published with the equal contribution of this author.

5.5.1. Purpose and scope

HuPSON provides a framework for: a) annotation of simulation experiments, b) text-mining based retrieval of information required for modeling, c) interoperability of algorithmic approaches used in biomedical simulation, d) comparability of simulation results and interoperability on different structural scales (human anatomy down to cells and molecules), and e) linking knowledge-based approaches

(such as ontologies) to simulation-based approaches (for instance, differential equation-based approaches).

5.5.2. Structural characterization

Structural features of HuPSON are summarized in Table 17. 1,067 (36%) of these classes were added manually whereas the other 64% of classes were integrated from related ontologies. Relatively high values of class number, leaves and width / average width together with a fanout factor of 0.71 are indicative of a broad coverage of the ontology, whereas the depth values of 10 (max.) and 5.5 (avg.) are indicators of a relatively good specificity of types to the domain.

Table 17. Structural parameters and their corresponding values calculated for HuPSON

	No. of roots	No. of classes	No. of synonyms	Max. depth	Average width	Tangled-ness	Fanout-ness
HuPSON	10	2920	7262	10	270	0.06	0.71

5.5.3. Functional characterization

Calculation of the system performance resulted in recall, precision and F-score of around 0.66 in the test set. Furthermore, participants from different working groups participating in the VPH Network of Excellence were asked to provide queries typical for the VPH domain. Table 18 shows that both ontology-based queries resulted in more true positive hits than their PubMed counterparts. These abstracts are considered to represent an “information gain” compared to the PubMed query results.

Table 18. Functional evaluation of HuPSON through answering competency questions as compared to the baseline PubMed search

Query expressed in free text	Hits of SCAIView query	Hits of PubMed query
Search the literature for fluid structure interaction models of the aneurysm simulating the pressure and its link to rupture	8 out of 9 abstracts TP	0 out of 0 abstracts TP
Find publications on velocity of blood flow and rupture outcomes of aneurysms	29 out of 59 abstracts TP	2 out of 3 abstracts TP

5.5.4. Application scenario

In order to show the applicability of HuPSON to independent domains, we applied it to Alzheimer's disease by challenging the system to retrieve and semantically filter the published knowledge related to simulation and modeling within this domain. Structural imaging has been recently shown to be a valuable tool in differential diagnosis of most dementias. To identify studies reporting the application of image analysis models to the differential diagnosis of Alzheimer's using MRI, we used the MeSH terminology in conjunction with HuPSON and performed a query in the SCAIView environment. 18 out of the 23 retrieved abstracts were relevant to the query and correctly identified such studies. From these documents, we were able to extract what specific model types are used in the

query context (e.g. “network diffusion models” and “logistic regression models”). This kind of information can help model developers choose an appropriate model for their research. it should be pointed out, that the combination of HuPSON with the AD disease ontology (see section 5.7) has the potential to contribute largely to the work done in the VPH-DARE project (see http://www.vph-dare.eu) where blood and lymphatic circulation simulation plays a major role.

5.6. CTO-NDD: Clinical Trial Ontology for Neurodegenerative Diseases

For a complete disease modeling process, linking clinical outcomes to their underlying molecular events is essential. Gaining insight into the mode of action of NDD-specific drug targets that were successfully evaluated through clinical trials in human populations will help to decipher mechanistic details underlying those targets and to guide the modeling efforts for target ID in AD. NDD-CTO harmonizes clinical readouts and is therefore an essential resource to make clinical trial data interoperable.

5.6.1. Purpose and scope

Although clinical trials measure the outcomes of a particular treatment or a particular biomarker, the knowledge behind the molecular mechanism underlying the measured outcomes remains implicit or unknown. In fact, the challenge of linking clinical outcomes to their underlying molecular events has been long of interest to the scientific community. The purpose of developing

CTO-NDD was to organize the key concepts of clinical trials and to mine the literature for molecular events linked to clinical endpoints. Scope of the CTO-NDD is focused on NDDs and covers the following areas: Clinical Trial Readouts, Clinical Trial Study, Clinical Trial Measurement Units, Clinical Imagings, Clinical trial methods, Exploratory Clinical Trials, Multicentered Trials, Pilot Trials, Practical Clinical Trials, Preclinical Trials, The Randomized Blinded Trial.

5.6.2. Structural characterization

Structural features of CTO-NDD are summarized in Table 19.

Table 19. Results of structural evaluation for the CTO-NDD ontology

	No. of roots	No. of classes	No. of synonyms	No. of leaves	Average width	Tangled-ness	Fanout-ness
CTO-NDD	12	483	1451	373	80.6	0.51	0.77

The high number of classes and leaves together with high values for avg. width and the fanout factor point towards a broad coverage of the ontology, whereas the values for depth can be used to show specificity of the types to the domain.

5.6.3. Functional characterization

Precision, recall, and F-score were calculated as listed in Table 20.

Table 20. Text mining-based evaluation of the CTO-NDD functional parameters

	Precision	**Recall**	**F-score**
CTO-NDD	0.79	0.70	0.74

Moreover, quality of CTO-NDD was assessed against five expert questions (Table 21). The following table summarizes the competency questions as well as the percentage of knowledge-domain coverage by CTO-NDD (the number of relevant documents divided by all the retrieved documents multiplied by 100).

Table 21. Functional evaluation of CTO-NDD by answering five competency questions

Competency question by expert	**Percentage correct answer by CTO-NDD**
Return references about clinical trials that use rating scales to measure treatment outcomes in Alzheimer's disease.	78%
Return references that report the application of using MRI for measuring clinical outcom of patients with relapsing-remitting and progressive multiple sclerosis.	62%
Return safety and efficacy studies mentioning receptor targets of drugs used for treatment of Parkinson's disease that have been shown to be effective in double-blind clinical trials.	88%
Return references of phase 3 clinical trials for treatment of epileptic seizures that met their primary endpoint.	32%

Return the clinical readouts for trials aimed at treatment of Alzheimer and provide the values for readout measurements	75%

5.6.4. Application scenario

A knowledge base was created for five NDDs (Alzheimer's disease, Parkinson's disease, Multiple sclerosis, Epilepsy, Amyotrophic lateral sclerosis (ALS)). The purpose was to collect data from completed trial documentations and classify them under biomarker and therapeutic categories based on endpoints, entities/disease phenotypes, and results from publications. The statistical content of the knowledge base concepts are shown in Table 22.

Table 22. Summary of documentation statistics published on the results of completed trials for 5 NDDs

Disease	Total number of complete trial documentations	Number of Biomarker documentations	Number of therapeutic documentations
Alzheimer's disease	276	57	219
Parkinson's disease	317	8	309
Multiple sclerosis disease	237	21	216
Epilepsy	217	1	216
Amyotrophic lateral sclerosis (ALS) disease	54	3	51

In the case of AD, the majority of clinical trials have measured safety and pharmacokinetics of therapeutics. Since existing clinical trials are performed with the aim of either biomarker discovery or treatment assessment, the information extracted from trial databases were grouped into Biomarker or Therapeutic classes and deposited into the knowledgebase. Table 23 exemplifies two entries of clinical trials in the knowledgebase for biomarkers and therapeutics of AD. It should be noted that NDD-CTO can be extended to capture relevant concepts measured in observational or epidemiological trials.

Table 23. Examples of biomarker (B) and therapeutic (T) entries found in clinical trial databases. Endpoints as identified by the ontology concepts have been related to the results of each trial through literature mining.

B/T	Title	Endpoints	Entities/Disease phenotypes	Results from publications
B	**Anti-Oxidant Treatment of Alzheimer's disease**	Safety/ Efficacy study	-Mild to moderate Alzheimer's disease -Vitamin E + C + alpha-lipoic acid -Coenzyme Q (CoQ) -Beta42 and a-beta40 protein -Probable Alzheimer's disease -Memory impairment	Use of vitamin E and C supplements in combination was associated with reduced AD prevalence (adjusted odds ratio, 0.22; 95% confidence interval, 0.05-0.60) and incidence (adjusted hazard ratio, 0.36; 95% confidence interval, 0.09-0.99). (PMID: 14732624)
T	**Alzheimer's disease: therapeutic Potential of**	Efficacy Study	-Transdermal estradiol drug -Medroxyprogesterone drug	Women treated with estrogen demonstrated improved performance on a test of semantic

	Estrogen		-Placebo Patch drug - Estrogen - Cognition decline -17-ß-estradiol - Estrone - FSH - ApoE - Mild-to-moderate dementia	memory (Boston Naming Test) compared with subjects who received a placebo. Estrogen appeared to have a suppressive effect on the insulin-like growth factor (IGF) system such that plasma concentration of IGF binding protein-3 was significantly reduced and plasma levels of estradiol and IGF-I were negatively correlated during estrogen treatment. (PMID: 11524467)

CTO-NDD goes beyond describing the knowledge of clinical trial readouts such as: biological readouts, cognitive readouts, physical readouts, primary and secondary endpoint measures and also, describing concepts related to different study types performed during clinical trials. CTO-NDD captures numerical descriptions and metrics that are routinely used in clinical trial studies. It also described different methods used for predicting clinical trial outcomes.

5.7. ADO: Alzheimer's Disease Ontology

Placing the context specific to disease characteristics and mechanism at the heart of any disease-specific model is a substantial criterion for correct and precise representation of disease-specific pathological

processes. In the absence of Alzheimer's-specific semantic frameworks such as terminologies or ontologies, which could potentially support target identification through modeling the disease mechanism in this thesis, there was an apparent need to create such semantic resource.

ADO was developed during the course of the Master thesis of Ashutosh Malhotra who was supervised by me in the course of this work. ADO has been published and made openly accessible to the AD research community [211].

5.7.1. Scope and purpose

Presently, the main online repository of ontologies, BioPortal (http://bioportal. bioontology.org/), contains only one generalized disease ontology, namely Human Disease Ontology (http://bioportal.bioontology. org/ontologies/1009, Last accessed: 22-4-2012) that has been designed to link distant datasets through disease concepts. However, the broad coverage and the lack of depth in this ontology restrict its usage for specific disease domains such as AD. Extensive efforts have been undertaken to organize published knowledge related to AD in the form of AlzSWAN knowledge base (http://www.alzforum.org/), which is supposed to be the reference repository for AD-based information available on the web. Still, information density or scalability remains a challenge for AlzSWAN. A first draft of an ontology representing clinical features, treatment, risk factors, and other aspects of the current knowledge in the domain of AD was developed. The hierarchical structure of the

ontology thus can serve as a robust navigation tree for terminology integration

5.7.2. Architecture and contents

ADO covers a wide range of key concepts and aims to engineer knowledge specific to AD. Modeled heterogeneity in this semantic framework tries to touch all relevant concepts but the range varies from very general (main ontology classes) to more specific concepts (concepts located at the ontology bottom "leaves"). The main views (root concepts) in ADO covering aspects of AD knowledge domain include: "Clinical", "Non clinical", "Etiological" and "Molecular and cellular mechanism". Each of these superclasses has its own subclasses (Figure 32, left). Subclasses mentioned under "Clinical (thing)" cover concepts that have contributed significantly to our understanding of the pathology, diagnosis and possible treatment options. Concepts defined under each subclass tend to be more domain-specific. For example, concepts representing stages of the disease are mentioned under clinical concepts (Figure 32, right).

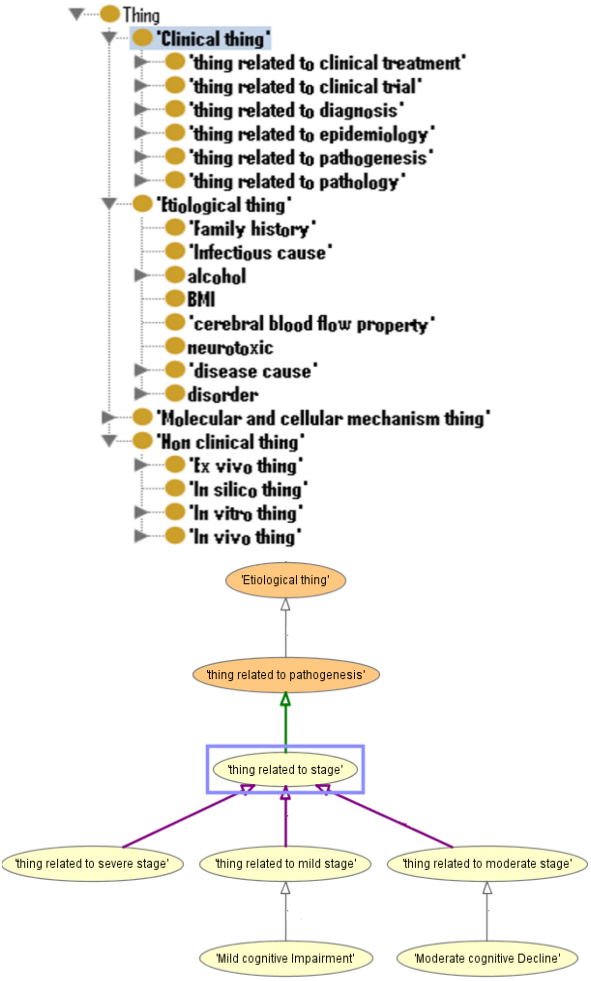


Figure 32. Representation of views incorporated in ADO (left) and example of disease stage concept and its subclasses (right). The left figure shows the hierarchical representation and the right figure illustrates the relational representation of the ontology concepts.

Currently much of the focus of AD research is devoted to preclinical studies that are conducted, typically on animals and has potential to play a vital role in drug discovery and development process. Concepts mapping all aspects of "Non-clinical" studies (*Ex-vivo*, *In-silico*, *In-vitro* and *In-vivo*) have also been incorporated into ADO, hence elaborating the knowledge related to animal models and bioassays used to better understand biological processes underlying Alzheimer's disease.

"Etiological view" forms the third root in ADO and it covers all aspects that might be responsible for the initiation of AD ranging from genetic factors and environmental influences to morphological changes whose effect varies from individual to individual. Clinical appearance of AD is also marked by anatomical changes as well as cellular and molecular cascades, which together manifest the neuropathological alterations observed in AD.

Furthermore, in ADO, the "Molecular and cellular mechanism" view is designed to cover all entities and biological mechanisms that find a possible role in AD. Studying behavior of neurons and other brain cells at the cellular and molecular level during AD can provide insights into the processes that mark its progression. The semantic relationships 'is a' and 'has a' and 'part of' were used to define relation types between pairs of concepts.

5.7.3. Structural characteristics

The structural features of the ontology reflecting topological and logical properties were measured by means of context-free metrics including depth and breadth (related to the cardinality of paths in a

graph), tangledness (related to multi-hierarchical nodes), and fan-outness (related to the dispersion of nodes). Table 24 shows various parameters, which were considered in the structural evaluation of ADO.

Table 24. Structural characterization of ADO

No. of roots	No. of classes	No. of synonyms	No. of leaves	Max. depth	Avg. width	Tangleness factor	Fan-out-ness
6	1486	2178	1221	12	302.91	0.16	0.82

The comparably high value of the Fan-outness factor exhibits the broad coverage of our ontology consisting of 1486 distinct classes. A tangledness factor of 0.16 is indicative for the presence of relatively few multi-hierarchical nodes in our hierarchy, which further confirms the coverage of the designed ontology.

5.7.4. Functional characteristics

The functional dimension of the ontology reflects the main purpose of that ontology by specifying a set of contextual assumptions about an area of interest. Functional evaluation measures how widely and precisely ontological concepts represent the semantic space for the indicated knowledge domain. The boundary of the knowledge domain addressed by ADO was estimated by calculating its fitness to an existing knowledge source (i.e. PubMed). Using our state-of-the-art text-mining environment, which takes ADO hierarchical structure and corresponding dictionary as input, we were able to evaluate ADO functionally on the prepared ‘test set’ (see Methods). As a

result, the functional metrics reported in Table 25 could be calculated.

Table 25. Results of the ontology evaluation using NLP-based approach

Test set corpus size	Precision	Recall	*F-score*
Independent test set of 200 abstracts	0.71	0.74	0.72

The result of this evaluation shows that the ontology in its current form can capture a wide range of AD concepts in the knowledge domain of AD scattered throughout scientific publications.

5.7.5. Expert evaluation

The expert panel's revision of the ontological "view" structure is considered as a genuine evaluation for disease ontologies. Following this, our ontology was manually curated by a clinician expert in the field who added certain clinically relevant concepts to ADO, increasing its pragmatic usability.

Two competency questions defined by the expert clinician and one from a pharmaceutical expert were selected to evaluate the semantic performance of ADO and its capability to return appropriate answers as following:

1. Return references linking amyloid beta to synaptic dysfunction in the mild stage of Alzheimer's disease.
2. Return references that associate the t-tau protein to frontotemporal dementia and Alzheimer's disease.

3. Return references containing clinical evidences that correlate CSF levels of p-tau with CSF Abeta 42 and cognitive decline.

To evaluate the above queries, we manually compared documents returned by ADO in SCAIView with documents returned by PubMed advanced search using the same queries. The results of this evaluation are summarized in Table 26.

Table 26. Results of competency questions evaluation using ADO within SCAIView compared to non-ontological search of PubMed.

Question No.	No. of SCAIView hits in total	Relevant SCAIView hits	No. of PubMed hits in total	Relevant PubMed hits
1	22 documents	19 documents	1 document	1 document
2	130 documents	14 documents	20 documents	7 documents
3	12 documents	out of top 20 7 documents	0 document	0 document

As shown in Table 26, querying SCAIView with the support of ADO in comparison to querying PubMed using comparable query formulations for all competency questions returns better results in terms of both sensitivity and specificity. These results indicate that ADO-supported information retrieval improves the chances for gaining better coverage with focused results in the same time as compared to naïve PubMed based searches. Also the knowledge gain in terms of concept recognition and enrichment is better when performing ontology based semantic search. To

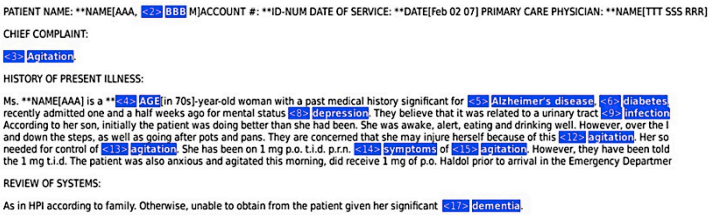
validate this, we used the third competency question and performed the query in SCAIView under two conditions: using the MeSH dictionary (without using ADO) and using ADO in conjunction to the NDD terminology. The reason is that the same corpus is indexed behind SCAIView so that the condition for queries remains fairly comparable. We then checked the document retrieval under both conditions and it turned out that the ontology based search using SCAIView (along with all MeSH concepts included as subset), was able to highlight approximately 5 times more entities than was possible using the MeSH dictionary alone. In this direction, we tried to derive inferences about the enrichment of the knowledge space surrounding the basic answer with the help of ADO concepts, e.g. how entities addressed in the competency questions are linked to other biological entities and processes (highlighted in the abstracts using ADO) in AD context.

Additional knowledge gained by means of ontological concept enrichment adds higher value to the original answer found by ADO. Although ontology based context modeling has been long acknowledged as a key aspect in a wide variety of problem domains, the full power of semantic search using a disease ontology comes with using a combination of the ontology with entity recognition, as shown by strategies discussed above.

5.7.6. Application scenario

Our motivation for development of ADO was to automatically extract domain specific knowledge related to AD, which can be used to gain further insights into the disease mechanism. To further

demonstrate the utility of ADO, we aimed at capturing knowledge beyond Medline abstracts. In an evaluation experiment independent from Medline, we used ADO to mine 650 AD-related electronic patient health records (EHRs) in order to systematically screen for other diseases or disorders that may exist simultaneously but independently in patients suffering from AD. In medical terms this concept is described as 'comorbidity analysis'. An example for the annotation of relevant terms in electronic patient records using the ADO dictionary is shown in Figure 33.


PATIENT NAME: **NAME[AAA, <2> BBB M]ACCOUNT #: **ID-NUM DATE OF SERVICE: **DATE[Feb 02 07] PRIMARY CARE PHYSICIAN: **NAME[TTT SSS RRR]

CHIEF COMPLAINT:

<3> Agitation.

HISTORY OF PRESENT ILLNESS:

Ms. **NAME[AAA] is a **<4> AGE[in 70s]-year-old woman with a past medical history significant for <5> Alzheimer's disease, <6> diabetes, recently admitted one and a half weeks ago for mental status <8> depression. They believe that it was related to a urinary tract <9> infection. According to her son, initially the patient was doing better than she had been. She was awake, alert, eating and drinking well. However, over the l and down the steps, as well as going after pots and pans. They are concerned that she may injure herself because of this <12> agitation. Her so needed for control of <13> agitation. She has been on 1 mg p.o. t.i.d. p.r.n. <14> symptoms of <15> agitation. However, they have been told the 1 mg t.i.d. The patient was also anxious and agitated this morning, did receive 1 mg of p.o. Haldol prior to arrival in the Emergency Departmer

REVIEW OF SYSTEMS:

As in HPI according to family. Otherwise, unable to obtain from the patient given her significant <17> dementia.

Figure 33. Comorbidity information captured in electronic health records using ADO. Risk factor concepts (comorbidities) included in ADO such as infection, and depression have been highlighted in text together with AD-related symptoms.

After analyzing the results obtained by calculation of term frequencies, we found that hypertension, diabetes and stroke are disorders that are reported to frequently occur in patients already suffering from AD. Furthermore, using our NER-driven literature mining machinery, we extracted a list of top 300 genes associated to these three diseases in the literature and compared it to the list of top

300 AD genes extracted from the literature using the same methodology. After curation for their relevance to the indication area, we found 19 common genes between AD and diabetes, 14 between AD and hypertension and 46 between AD and stroke (Figure 34).

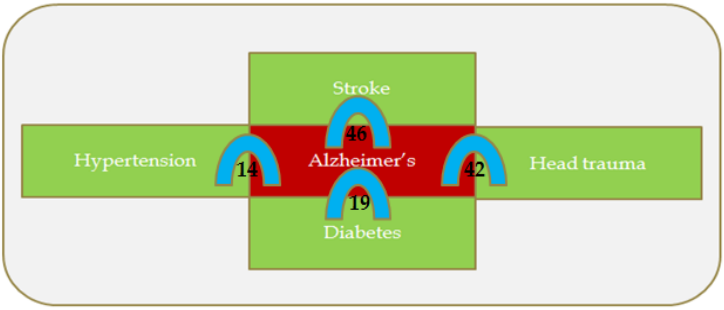


Figure 34. Four major comorbidities frequently mentioned with AD in EHRs and number of genes reported in the literature to be commonly involved in Alzheimer's and these comorbid diseases. A higher number of genes were retrieved for the overlap of AD with stroke (46) and head traumas (42) at the mechanistic level than diabetes (19) and hypertension (14).

5.7.7. Perspective and future outlook

With the public release of ADO (http://bioportal.bioontology.org/ontologies/ ADO), the hope is to reach out to the broader AD research community and to continuously improve the quality of the ontology. Furthermore, the project team will continue to review articles, abstracts, and other resources relevant for the

domain and to update and maintain ADO. A potential application of ADO in the future would be improvement of AD-specific models through semi-automated model updating process using ADO-driven text-mining systems. Moreover, ADO can be used to construct BEL models of AD physiopathological aspects.

CHAPTER 6. Integrative Modeling of Pathological Components associated to Alzheimer's Disease

In this chapter, I describe the modeling activities related to the AD pathology, in response to the key scientific questions of this thesis:

- "are different pathways leading to the manifestation of AD?",
- "are plaques and tangles secondary events or downstream effects?", or
- "what role is the dimension of time playing in the pathophysiology of AD?".

As discussed in Chapter 2, the putative mechanisms underlying the AD pathology have been formulated in the form of three major hypotheses, including amyloidosis, inflammation, and infection. From the systems point of view, these pieces of information need to come together and generate a bigger picture of the AD mechanism. The amyloidosis hypothesis describes the dysfunction of the CNS most important organ, the brain, whereas inflammation and infection both form the two sides of the same coin of the immune system deregulation. A third perturbed component of the system, which has not yet been explicitly put forward as a hypothesis, is the endocrine system (what I call "hormone hypothesis"). Interestingly, this

component came into the picture after an unexpected observation of enriched hormone proteins and their receptors in the literature-derived data. The importance of the endocrine component is that the nervous and immune systems communicate via multiple hormonal routes and mechanisms and their interactions provide a finely tuned regulatory system, which is disturbed under the disease condition.

6.1. Exploratory data analysis

Exploratory data analysis is an analysis approach – often with visual methods - that focuses on summarizing the main characteristics of data by identifying general patterns in the data and features of the data that might not have been anticipated. In fact, the purpose of performing exploratory data analysis is to find the right formulation of the scientific question or hypothesis. In contrast to confirmatory analysis, which involves testing a previously established hypothesis, exploratory data analysis involves a broad investigation to explore datasets for getting insights into potential new hypotheses. This approach, therefore, has important implications for analysis of large and high-throughput biological data; for example, biological networks have been used for exploratory data analysis to discover modular structures in high-throughput data [212] or pathway diagrams have been constructed to facilitate exploratory analysis and to enhance biological discovery in large datasets [213].

Accordingly, to explore directions towards answering the scientific questions in this thesis and to obtain the very first intuition about the available data sources, data qualities and their information value,

both literature-derived and database-derived data, after passing through expert curation, were subjected to an initial analysis in order to get a preliminary insight into their knowledge content and how best this knowledge can be used in the modeling process. This section reports on the feasibility study on the NDD- and AD-related data and knowledge.

6.1.1. Molecular overlap between NDDs

It is well known that NDDs share common disease features [214]. Thus, an entry point to understanding the molecular etiology of NDDs is analysis of existing knowledge about genes and proteins that they share.

A spectrum of six NDDs - presenting 'dementia' as a common symptom – was surveyed in the literature using dictionaries for the Human Genes/Proteins and MeSH terms in the SCAIView literature mining environment, and finally, a list of genes/proteins associated to these NDDs was generated. For each disease, the output list was manually curated to ensure the association between the gene and the disease of interest. The overlap of genes was computed between pairs of diseases and presented as a matrix (Figure 35).

	Huntington	Lewy body	Pick	Creutzfeldt -Jakob	Alzheimer	Parkinson
Huntington		58	39	25	124	100
Lewy body			58	31	180	155
Pick				27	124	80
Creutzfeldt-Jakob					72	43
Alzheimer						350
Parkinson						

Figure 35. Matrix containing number of overlapping genes between six NDDs.

The results clearly show that AD shares a significant number of genes with Parkinson's disease (350 genes), followed by Lewy body disease (180 genes) and Parkinson's disease (155 genes). This implies that there exists a similar molecular etiology underlying these NDDs. The top-ranked common genes among these three diseases include APP, APOE, MAPT, SNCA, SNCB, CHAT and PRNP.

6.1.2. Potential biomarkers of AD

In general, there are two categories of biomarkers: "known valid biomarkers", which have been measured experimentally with a well-established performance and "probable valid biomarkers", which have been measured in an analytic test system but more evidence on their performance is needed to be approved [215]. The second category represents a vast majority of proposed biomarkers in the literature (potential biomarkers), which report a measured indication under the disease conditions. This information can be used in guiding the pathway analysis as I propose in the next chapter.

In order to obtain a list of potential biomarkers, the following query was performed with the help of SCAIView and the biomarker terminology:

(([MeSH Disease:"Alzheimer Disease"]) AND [BioMarker Terminology Node:"Evidence Marker"])

The long list of retrieved potential AD biomarkers - comprising of 1009 genes/proteins - was filtered for expression evidence and was subjected to manual inspection of sentences. Figure 36 depicts the workflow for identification of these markers. Finally, 366 proteins reported to potentially act as disease biomarkers were identified and underwent the process of pathway enrichment analysis.

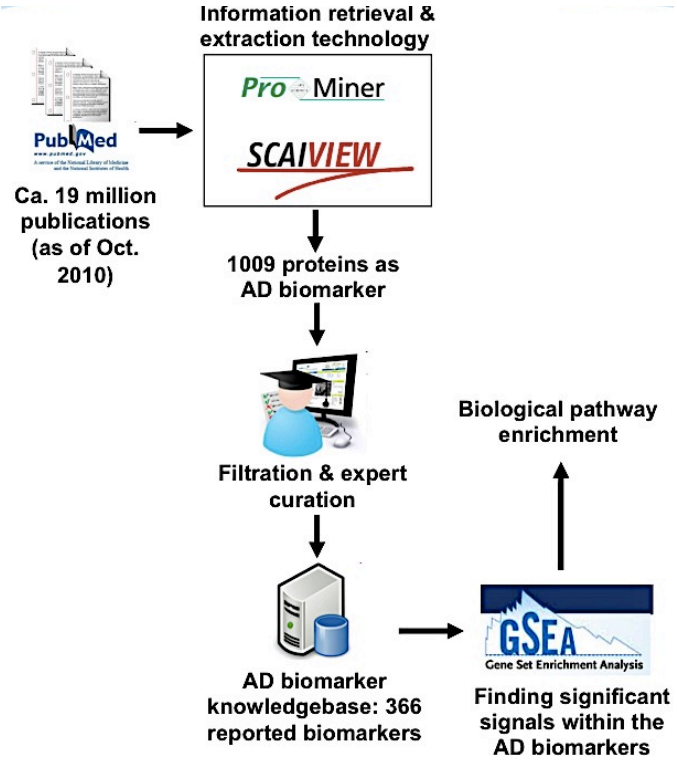


Figure 36. Diagrammatic workflow used for identification of potential AD biomarkers in literature. The biomarker terminology was used to query for potential expressed biomarkers of AD in PubMed, which were subjected to filtration, manual curation and pathway enrichment analysis.

Table 27 represents the top significant pathways that are enriched with potential biomarkers of AD. This exploratory pathway analysis indicates that experimentally measured indicators of AD initially

point to inflammatory and cell death processes under disease condition.

Table 27. Results of pathway enrichment analysis for literature-derived list of potential AD biomarkers

Pathway	p-value
Amyotrophic lateral sclerosis (ALS)	2.78E-13
HIV-I Nef: negative effector of Fas and TNF	1.52E-09
Apoptosis	1.11E-07
Antigen Dependent B Cell Activation	1.14E-07
Apoptotic Signaling in Response to DNA Damage	3.71E-07
Keratinocyte Differentiation	6.41E-07
Cytokines and Inflammatory Response	8.74E-07
Prion diseases	9.81E-07
Stress Induction of HSP Regulation	1.24E-06
FAS signaling pathway (CD95)	1.30E-06
Cytokine Network	2.68E-06
Alzheimer's disease	2.99E-06

6.1.3. Target receptors of the human brain

Cell surface receptors and ion channels constitute targets for one-third of the current medicinal drugs with G-Protein-Coupled Receptors (GPCRs) forming the largest group of drug targets (36% of the human targets) [216]. The International Union of Basic and Clinical Pharmacology Committee on Receptor Nomenclature and

Drug Classification (NC-IUPHAR) has released guidelines for nomenclature and classifying these targets. They also provide the IUPHAR-DB database containing peer-reviewed pharmacological, chemical, genetic, functional and anatomical information on receptor and channel targets for human, rat, and mouse (http://www.iuphar-db.org).

Using the annotation fields of the IUPHAR-DB, I was able to download all the GPCR and ion channel receptors that were annotated to be expressed in the human brain tissue. After the manual check, 69 GPCRs and 4 ion channel receptors were identified to specifically express in the human brain, some of which have been already targeted by approved drugs (Table 28).

Table 28. List of curated receptors that are expressed specifically in the human brain

GPCRs	KiSS1, MCH2, 5-HT4, RXFP4, PKR2, P2Y6, 5-HT1B, FZD2, MCH1, A2A, H2, GPR116, RAIG1, α 2C-adrenoceptor, RAIG2, A1, H1, 5-ht5a, PKR1, P2Y11, 5-HT1F, NPBW1, NPFF2, NPFF1, GPR125, P2Y12, P 2Y2, α2B-adrenoceptor, FZD1, GPRC6, sst4, RXFP2, A3, NTS 1, P2Y14, P2Y13, NOP, mGlu3, ETB, GPR158, GPR98, ghrelin, 5-HT2B, 5-HT7, mGlu2, MT1, GLP-1, BB2, ETA, H4, H3, LH, GABAB2, GPR126, GPR179, A2B, α1A-adrenoceptor, DP2, P2Y4, α2A-adrenoceptor, GPR56, FFA1, RXFP3, P2Y1, TP, RX FP1, EP4, VPAC1, GPBA
Ion channel receptors	5-HT3C, ZAC, 5-HT3D, 5-HT3E

6.1.4. Human brain interactome

Given the complexity of the human brain structure and in the absence of brain-specific molecular interaction models, it seemed reasonable to build an interaction model that represents a good resolution of protein interactions at the level of brain regions or tissue types. Tissue specificity is an important aspect of many diseases, which is frequently ignored when disease process models are built. Knowing that proteins and pathways play different roles in diverse tissues and cell types, functionality and functional relationships of molecular interactions should be resolved when modeling disease processes for target ID is performed. The rationale is that disease phenotypes are caused by tissue-specific pathology and thus, building tissue-specific functional networks can support prioritization of disease genes and association of disease phenotypes to underlying molecular interactions [217]. As demonstrated in Chapters 8 and 9, the brain interactome has been applied to modeling brain-specific disease pathways that are associated to the damaged areas of the brain anatomy. Application of the brain interactome goes beyond disease modeling by mapping the pharmacological space of CNS drugs to this network and including drug-target and drug-drug interactions on top of that. Such a model – called as the Human Brain Pharmacome - can be constructed with the aim of polypharmacology analysis or feasibility study of drug repurposing.

In order to construct a context-sensitive network model representing possible protein interactions in the human brain - regardless of normal or disease conditions - in a tissue-specific manner, a brain-

specific PPI network representing 15 brain regions was reconstructed using the global human protein interaction network [218]. These regions have been annotated to the interactions using co-expression profiles and include amygdala, cerebellum peduncles, cingulate cortex, hypothalamus, medulla oblongata, occipital lobe, parietal lobe, pons, prefrontal cortex, subthalamic nucleus, temporal lobe, thalamus, caudate nucleus, cerebellum, and globus pallidus. Presence or absence of a particular interaction in a brain region has been indicated by numbers so that number 1 indicates the presence of that interaction in that particular brain region and number 0 indicates the absence of a particular interaction in a particular brain region. It appears that some interactions occur in several regions and some others are region-specific.

The brain interactome has been constructed based on the PPI information from 21 databases and is composed of 7255 proteins and 47882 interactions. Edges are manually annotated and supported by additional attributes including source interaction databases, supporting experimental methods and evidence, evidence sentence from the literature, and source PMIDs in the literature (if available). Figure 37 illustrates an example snapshot of edge annotations in the brain interactome, which has been visualized in the Cytoscape environment.

ID	Amygdala	CerebellumPeduncles	CingulateCortex	Hypothalamus	Databases	Direct_evidence	Supporting_evidence	PMIDs
CTSD (pp) IGFBP3	1	1	0	1	HPRD|	in_vitro|		9275067|
CHGB (pp) UBTF	1	1	1	1	HOMOMINT|HPRD|INTACT|MDC...	physical_interaction|sd4-t...		16169070|
FGB (pp) CALR	1	1	1	1	BIOGRID|BIOVERSE|CCSB-HPR...	biophysical|immunoprecip...		7592883|
PSMD3 (pp) ATG12	1	1	1	1	INTNET|ORTHO-noncore|	physical_interaction|	co_expression|domain_domai...	
ACTN2 (pp) RAVER1	0	0	0	0	BIOGRID|HPRD|UNILEVER|	in_vivo|two_hybrid|		11724819|
MAPK3 (pp) PTPRE	1	1	1	1	HPRD|	in_vitro|		12754301|
DIMT1L (pp) BYSL	1	1	1	1	COCIT-ORTHOLOGY|INTNET|O...	physical_interaction|	co_expression|domain_domai...	
MYOD1 (pp) HEY1	0	0	0	0	BIOGRID|CCSB-HPRD|HPRD|OP...	in_vitro|in_vivo|		11279181|
MAPK3 (pp) RPS6KA1	1	1	1	1	BIOGRID|BIOVERSE|CCSB-COCI...	in_vitro|in_vivo|	cocitation|interolog-worm|	10965886|
RBPJ (pp) RRMS1	1	1	1	1	HPRD|UNILEVER|	in_vivo|		15451426|
GSK3B (pp) GYS1	1	1	1	1	BIOVERSE|HPRD|INTACT|UNILEV...	in_vitro|phosphorylation|p...		11427888|
DDX24 (pp) RPS24	1	1	1	1	INTNET|	physical_interaction|	co_expression|domain_domai...	
VASP (pp) ACTA1	0	0	0	1	HPRD|UNILEVER|	in_vitro|two_hybrid|		10438535|
BATF3 (pp) ATF3	1	1	1	1	HPRD|OPHID|UNILEVER|	in_vitro|		12805554|
POLR3H (pp) SNAPC3	1	1	1	1	COCIT-REACTOME|REACTOME|...	indirect_complex|neighbo...		
RNF4 (pp) ESR1	1	1	1	1	BIOVERSE|HPRD|OPHID|UNILEVER|	in_vitro|in_vivo|		11696545|
RFC2 (pp) TRRAP	1	0	1	1	INTNET|	physical_interaction|	co_expression|domain_domai...	
IGFBP3 (pp) KLKB1	0	1	0	1	BIOGRID|HPRD|OPHID|UNILEVER|	in_vivo|		10226805|
PARP1 (pp) TOP1	1	1	1	1	CORUM|	literature_annotated_com...		
CA12 (pp) SCSM2	1	1	1	1	HPRD|MDC|UNILEVER|	lacz-4-two_hybrid|two_hy...		16169070|
HIF1A (pp) TP53	0	1	0	1	BIND|BIOGRID|CCSB-curated|CO...	affinitycapture_ms|immun...		11593383|
CBL (pp) EPHA2	1	0	1	1	BIOVERSE|HPRD|UNILEVER|	in_vivo|		12496371|
ADORA2A (pp) ADA	0	0	0	1	BIOGRID|BIOVERSE|CCSB-curate...	affinitycapture_ms|in_vivo|...	share_go_process|	11125033|
RPS6 (pp) VPS39	1	1	1	1	INTNET|	physical_interaction|	co_expression|domain_domai...	
DAPK3 (pp) PAWR	1	0	0	1	HPRD|INTACT|UNILEVER|	coip|in_vitro|in_vivo|physic...		10602480|
RPS21 (pp) RPL14	1	1	1	1	COCIT-REACTOME|CORUM|INT...	"indirect_complex|literatur...	co_expression|domain_domai...	10659855|
POP5 (pp) RPP25	1	1	1	1	BIOGRID|CORUM|HPRD|INTACT|...	in_vitro|literature_annotat...		15096576|
VPS18 (pp) PTPRA	1	1	1	1	COCIT-ORTHOLOGY|HPRD|OPH...	in_vivo|	interolog-yeast|	11382755|
IKBKAP (pp) GNB2L1	1	1	1	1	INTNET|	physical_interaction|	co_expression|domain_domai...	
ARHGEF7 (pp) SHAN...	0	0	0	0	BIOGRID|BIOVERSE|HOMOMINT-...	in_vitro|in_vivo|two_hybrid|	interolog-mus_musculus|	12626503|
CCNH (pp) GTF2H4	0	1	1	1	CORUM|	literature_annotated_com...		

Figure 37. Visualization of annotated attributes in the brain interactome in Cytoscape environment. Protein-protein interactions are coded for their presence (1) or absence (0) in particular brain regions based on their co-expression and further substantiated with additional evidence and information from databases and publications.

Since the brain interactome captures a large amount of PPI information in the form of a hairball network with a giant component, it is best suited for "top-down" analysis methods. Indeed, the attributes embedded in the annotation space of the brain interactome will be used to support such a top-down analysis by providing the possibility to filter down the network and create various facets of the interactome model based on brain regions, e.g. generation of subnetworks representing interactions occurring only in hypothalamus or in cingulate cortex or both.

It should be noted that the quality of this network model can be improved as more data become available on protein-protein interactions. For example, the majority of this interactome is based on retrieved data from protein and literature databases, which change and update dynamically. Therefore, an automatic survey of new data

releases or newly published literature can facilitate regular update of both interactions and attributes of the network.

6.2. Modeling of molecular processes underlying AD

As a first step towards exploratory analysis of disease processes underlying AD at the molecular level, the first version of the PPI model representing amyloid hypothesis in AD was constructed. The purpose was to explore the knowledge space published around molecular mechanisms underlying AD, visualize this knowledge space as a network model, and analyze it for the kind of pathways that are over-represented or enriched in the pathology of AD.

For this purpose, the literature in SCAIView were queried with the "MeSH: Alzheimer's disease" key word and searching for the list of "Human Genes/Proteins" associated to AD. In order to provide the relevant set of proteins for construction of the network (known as 'seed proteins'), 35 AD susceptibility genes were acquired from the AlzGene database [219], which is considered to be the gold standard for AD-related genes. Out of 35 genes, 31 genes were mapped onto the genes retrieved from the literature. These seed proteins were used as input to the BIANA tool and after manual curation, a network of 684 nodes and 893 edges was constructed (Figure 38).

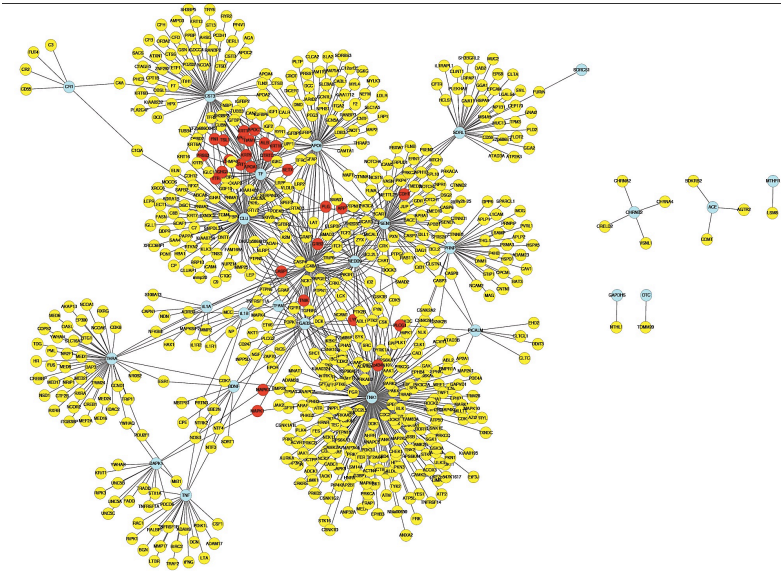


Figure 38. A first AD-specific PPI network model constructed around AD genes from AlzGene database. Nodes in cyan represent hubs and nodes in red represent linker (inter-module) proteins.

When submitted to the DAVID tool for functional annotation [220], the genes of this network showed significant enrichment for 72 KEGG pathways. To obtain the modular distribution of pathways, the hits were sorted based on the gene memberships in annotated pathways (Figure 39). These results show that, in this distribution, intracellular signaling pathways that are involved in the cell cycle and cell proliferation processes (reflected in the cancer and MAPK signaling pathways) are dominant. But the interesting observation is the presence of immune system-related pathways, i.e. chemokine

signaling pathway as well as natural killer cell signaling. Another important observation is the prominent involvement of a hormone-dependent signaling pathway, namely the insulin-signaling pathway.

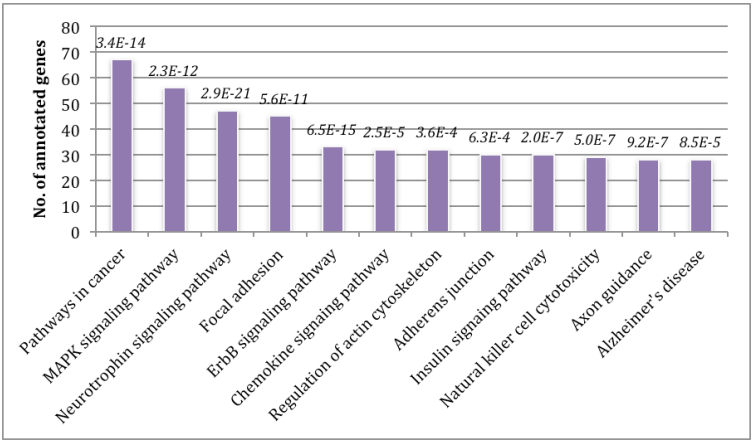


Figure 39. Results of functional annotation of genes in the AD network model to KEGG pathways. Pathways in cancer and MAPK pathway are the predominant pathways in this AD model and may point to the dysregulation of immense intracellular signaling and perturbation of cell cycle. Alzheimer's pathway also emerges as a part of the bigger picture and indicates the relevance of the model.

Taken all the above observations together, the results of exploratory data analysis - as the very first attempt at making sense of available data - highlighted the involvement of 'immune' and 'hormone' signaling components in the core pathological processes related to AD pathway. It should be noted that because the model generated here is an initial draft for obtaining directions and insights into the

disease mechanism, its coverage is low and thus, the signal is weak. Although these results may not show a clear pattern relevant to disease mechanism and the signals are weak, they at least provide a hint and further direction for exploring relevant hypotheses, for example, on the role of the immune and hormone components in pathology of AD. In the next chapters, the role of immune and hormone signaling pathways in the AD pathology is investigated more concretely.

CHAPTER 7. Modeling of the Immune System Component associated to Alzheimer's Disease

7.1. Introduction

Neurodegenerative dementia is the main clinical manifestation in a multitude of diseases such as Alzheimer's, Lewy bodies, Huntington's and Parkinson's diseases. It is often a feature of other pathological conditions including Creutzfeldt-Jakob disease as well as a couple of neurodegenerative diseases and syndromes. Moreover, several arrays of evidence suggest that infectious agents can induce the pathogenesis of dementia in infected people [221]. Among these infectious agents, viruses have been hypothesized to play a role in the development of dementia and dementia praecox [222],[223].

It has been shown that HIV infection can cause dementia in at least 20% of all HIV infected patients [224],[225] and it appears that the AIDS dementia complex is confined to the later phases of systemic infection by the human immunodeficiency virus [226]. Jang *et al.* (2009) showed that mice infected by the highly pathogenic H5N1 influenza also manifest neuropathological signs and they suggested that any neurotropic influenza virus that activates the immune system in the brain can potentially lead to Parkinson's and Alzheimer's diseases [227].

The common observation among virus-induced dementia cases is the prolonged inflammatory responses as well as long latency period [228]. Interestingly, this time-dependent progression of the disease also occurs in age-related neurodegenerative diseases including Alzheimer's. For example it has been proposed that long-term accumulation of insults to the immune system leads to the persistent activation of microglial cells which mediate progressive neurotoxicity [229]. Indeed, several lines of evidence in animal models indicate that exposure to bacterial lipopolysaccharide (LPS) or selective neurotoxins induces microglial activation which can last for years and cause the progressive and cumulative loss of neurons over time [230],[231],[232]. Jang *et al.* observed a similar trend in H5N1 infected mice whose CNS showed persistent microgliosis and prolonged inflammatory response even after the infection had resolved. Although inflammatory responses accompanied by over-activation of the immune system are hallmarks of neurodegenerative diseases [233], neither non-steroidal anti-inflammatory drugs (NSAIDs) nor highly active antiretroviral therapy (HAART) have led to a definitive therapeutic outcome [234],[235].

On the other hand, emerging evidence suggest that the current amyloid hypothesis may not capture the causative mechanism underlying the Alzheimer's dementia and a rethinking of the relationship between amyloid aggregation and neurotoxicity is required. Further support to this comes from the recent failures of anti-amyloid drugs including 11 investigational drugs for Alzheimer's disease [236]. All these have triggered the idea that the

popular amyloid hypothesis may not well explain the cause of the disease or even could be a secondary event [237].

HIV-associated dementia, among the virus-induced dementia types, is the best-studied model, which may provide some evidence on existing alternate pathogenic pathways in dementia. To gain an insight into the disease mechanism underlying AIDS dementia complex, we adopt a knowledge-based analysis approach by integrating literature data and biological networks.

7.2. Methods

7.2.1. Text mining and information retrieval

We employed SCAIView and accordingly a sub-corpus of the MEDLINE abstracts related to the keywords "infection AND neurodegeneration", and "HIV AND neurodegeneration" were defined and queried for the instances of the human genes/proteins. The first query is rather general and covers neurodegenerative effects of both viral and non-viral infectious agents (including microbes and fungi) whereas the second one retrieves entities, which are involved in HIV-associated neurodegeneration. The output for each query was a list of human Genes/Proteins named entities extracted from PubMed abstracts and ranked based on relative entropy measurement in descending order. 315 genes/proteins were retrieved for the first query and 143 genes/proteins for the second query. For each entity, the retrieved abstracts were manually checked for the correct relevance of the entity to the disease context and possible false

positives were removed. This led to a reduction of the number of genes and proteins to 170 and 88 genes/proteins, respectively. A third sub-corpus of abstracts using the keywords "virus AND infection" was defined and the results were compared to the human-HIV interacting proteins from the VirusMINT database [238].

7.2.2. Calculation of overlapping probabilities

When comparing two top ranking gene lists obtained from two different queries, the overlapping gene set should be tested for the probability of this being a chance event. We used the hypergeometric probability approximated by Fisher's exact test as suggested by Fury *et al.* [239] to calculate the p-values for overlaps between gene lists (see Results).

7.2.3. Data acquisition

In order to check for the relevance of the results, all human proteins that interact with HIV proteins were downloaded from the VirusMINT database. After removing the viral proteins and duplicates, 173 proteins were obtained and normalized on their corresponding gene official symbols. Incipient (MCI) and acute Alzheimer's disease (AD) gene expression data were obtained from the study of Blalock *et al.* [240]. All the three datasets (i.e. infected human proteins and expressed genes under MCI/AD) were compared to the Immunome database [241] and the common signatures between these datasets and the Immunome-derived dataset were extracted.

7.2.4. Interaction network reconstruction

We mapped our gene list from the query "HIV AND neurodegeneration" onto the brain interactome and obtained a network with 85 nodes and 58 edges. The network consists of a connected giant component with 37 nodes that bears the maximum number of total edges (54 out of 58). In order to build a dementia network model specific to human on a relevant and scientific basis, an OMIM-based dementia network was reconstructed by extracting the following descriptions from the OMIM Morbid Map [242] and pooling their corresponding genes into one union list after gene name normalization: dementia, Alzheimer, Parkinson, Lewy body, Huntington, and Creutzfeldt. These genes were fed into our text-mining system and only 40 of them were found to have 2 or more reference publications in PubMed. We used these 40 genes as "seed genes" to reconstruct the protein-protein interaction network in the BIANA environment. The BIANA tool retrieves all possible interactions among these proteins and one or two levels of other neighboring proteins from a variety of databases and creates a network model around the specified seed genes. This network contains 886 nodes and 1264 edges.

7.3. Results

7.3.1. Distinction between infection and neurodegeneration

Dealing with virus-induced dementia raises the important question whether the current knowledge space surrounding AIDS-dementia

complex represents genes/proteins involved in HIV pathogenesis or the subsequent neurodegeneration. Comparison between the results of Q2 and human proteins interacting with HIV during the course of viral infection (data from VirusMINT database) showed only 7 genes/proteins in common. This indicates that our query results are relevant to the neurodegeneration rather than the infection process. Gene list from Q2 was annotated by MSigDB tool (www.broadinstitute.org/msigdb) and it was found that this list is significantly enriched for the apoptosis, inflammatory pathways, and neurodegenerative diseases (Table 29).

Table 29. Annotation results for the 88 genes obtained from the sub-corpus of HIV neurodegeneration

Geneset Name	Description	No. of genes	p-value
Apoptosis pathway	-	9	1.52E-08
IL-10 pathway	The cytokine IL-10 inhibits the inflammatory response by macrophages via activation of heme oxygenase 1.	6	2.15E-08
Chemical pathway	DNA damage promotes Bid cleavage, which stimulates mitochondrial cytochrome c release and consequent caspase activation, resulting in apoptosis.	7	2.83E-08
Neurodegenerative diseases	Genes involved in neurodegenerative diseases	8	9.58E-08
HSP-27 pathway	Hsp27 oligomers have molecular chaperone activity and protect heat-stressed cells against apoptosis.	6	9.60E-08
P53 pathway	p53 induces cell cycle arrest or apoptosis under conditions of DNA damage.	6	9.60E-08
CCR5 pathway	CCR5 is a G-protein coupled receptor expressed in macrophages that recognizes chemokine ligands and is targeted by the HIV envelope protein GP120.	6	2.16E-07

HIV-NEF pathway	HIV-infected CD4 helper T cells may express Fas ligand, which binds to the Fas receptors of uninfected cells and induces apoptosis.	9	2.31E-07
IL1R pathway	The cytokine IL-1 stimulates its primary receptor, IL-1R1, which induces transcription of inflammation-related genes such as interferons.	7	6.01E-07
Mitochondria pathway	Pro-apoptotic signaling induces mitochondria to release cytochrome c, which stimulates Apaf-1 to activate caspase 9.	6	6.03E-07

7.3.2. Relevance to neurodegeneration and dementia

Currently there is no clear approach to measuring the contextual quality of the genes mentioned in the body of literature. For our purpose, we constructed a gene-literature bipartite network (out of the results of Q2) in which the genes and PMID of their reference publications are connected. The Eigenfactor value for each PMID was included as the weight of edges. Eigenfactor is a scoring metrics, which ranks scholarly journals according to the local citation information as well as considering the entire network of citations. Filtering out edges with the Eigenfactor values below 0.1 left a network of 45 genes, which are cited in the quality publications (Figure 40). These genes were submitted to the MSigDB for pathway enrichment analysis; the results indicate a significant enrichment for neurodegenerative and apoptotic pathways (Table 30).

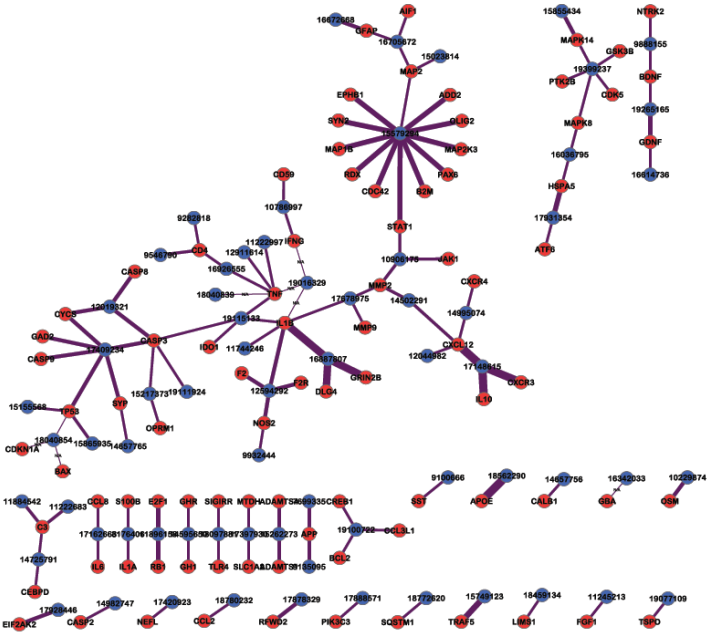


Figure 40. Gene-literature bipartite graph. Nodes in red and blue represent genes and their paper references respectively. A connection between the red and blue nodes is made if the gene is referenced in the corresponding publication.

Table 30. Pathway enrichment analysis of the 45 genes at the core of gene-literature bipartite network

Geneset Name	Description	No. of genes	p-value
Neurodegenerative diseases	Genes involved in neurodegenerative diseases	6	4.60E-07
Chemical pathway	DNA damage promotes Bid cleavage, which stimulates mitochondrial cytochrome c release and consequent caspase activation, resulting in apoptosis.	5	6.71E-07
Apoptosis pathway	Genes involved in apoptosis.	6	9.79E-07

Pancreatic cancer	Genes involved in pancreatic cancer:	7	1.52E-06
Alzheimer's disease	Genes involved in Alzheimer's disease	5	2.40E-06
D4GDI pathway	D4-GDI inhibits the pro-apoptotic Rho GTPases and is cleaved by caspase-3.	4	2.59E-06
IL-10 pathway	The cytokine IL-10 inhibits the inflammatory response by macrophages via activation of heme oxygenase 1.	4	2.59E-06
HIV-NEF pathway	HIV-infected CD4 helper T cells may express Fas ligand, which binds to the Fas receptors of uninfected cells and induces apoptosis.	6	5.84E-06
HSP-27 pathway	Hsp27 oligomers have molecular chaperone activity and protect heat-stressed cells against apoptosis.	4	6.45E-06

To realize the extent of relevance the viral neurodegeneration might have to dementia disorders, we compared the gene lists that we had obtained from querying PubMed abstracts for Alzheimer's dementia and Parkinson's dementia with the gene list from Q2. Both gene-lists showed high overlaps with the Q2 results (68.18% and 43%, respectively) that did not match by chance (hypergeometric p-values=5.0090e-13, and 9.7188e-22, respectively). Gene set enrichment analysis was performed on these overlapped gene sets for pathway annotation and for both overlapping sets similar pathway annotations were retrieved that include apoptosis, Alzheimer's disease, and IL1 pathway.

7.3.3. Relevance to the inflammation and immune system activation

In the pathway annotation of the 88-gene list obtained from the HIV neurodegeneration query, a predominant presence of inflammatory

pathway elements was observed with the IL10 pathway as the second most significant category. It is being discussed that inflammation is the hallmark of CNS disorders and its link to the activation of the immune system is established. To investigate the extent of relevance between neurodegeneration and the inflammatory induced immune response, we first determined which pro- and anti-inflammatory cytokines are mentioned in the literature to be active in the brain; using two keywords namely, "proinflammatory AND brain" and "anti-inflammatory AND brain", the literature corpora were queried in SCAIView for those genes/proteins involved in the brain pro-/anti-inflammatory processes. Both pro- and anti-inflammatory genes in the "brain" were identified by comparing the results of SCAIView queries against the Cytokina Database (Table 31). To the best of our knowledge, this is the first curated list of inflammatory genes/proteins reported in the human brain.

Table 31. The list of brain-expressed pro- and anti-inflammatory genes collected from the literature

Pro-inflammatory	CCL1, CCL17, CCL2, CCL20,CCL27, CCL3, CCL4, CCL5, CCL7, CCL8, CD40LG, CSF2, CSF3, CX3CL1, CXCL1, CXCL10, CXCL12, CXCL2, CXCL5, CXCL9, FASLG, IFNA1, IFNB1, IFNG, IL12A, IL15, IL17B, IL18, IL1A, IL1B, IL2, IL23A, IL6, IL8, LIF, LTA, OSM, TGFB1, TNF, TNFSF12
Anti-inflammatory	CSF3, EPO, IFNB1, IL10, IL11, IL13, IL4, TGFB1

Then, the MCI and AD expression data was used to see the activity range of these genes under disease conditions. After projection of expressions onto these inflammatory genes it was observed that none of the 'anti-inflammatory' genes was found amongst the MCI or AD over-expressed genes; however from 'pro-inflammatory' genes, IFNG and IL18 in mild cognitive impairment, and CXCL2, IL1A, IFNG, IL15, and TGFB1 in Alzheimer's disease were found to become upregulated. These results imply that the homeostatic balance between pro- and anti-inflammatory pathways might be perturbed under neurodegenerative conditions.

7.3.4. Pathway-pathway interaction map

In order to get a vision on the structure of the crosstalking pathways involved in the HIV-induced dementia, our gene list from the "HIV AND neurodegeneration" query was mapped onto the human protein interactome network as described in the Methods section. The giant component of the network was subjected to the pathway analysis by submitting its constituent genes to the MSigDB database. It is observed that the network is enriched for DNA damage pathways, apoptotic pathways, p53 pathway and neurodegenerative diseases pathways. The confirmation for relevance of the giant component to neurodegenerative diseases comes from the significant overlap of this component with our OMIM-based dementia network (55.88%, hypergeometric p-value= 1.2249e-14).

Visualization of the interconnections among pathways in the giant component was performed in the Cytoscape environment using the PNmerger plugin [243]. This plugin annotates the network proteins

with the KEGG pathway information, extends the network to the full pathway membership and also presents the potential crosstalk elements among pathways (Figure 41). Identified pathways and their crosstalk partners are summarized in the Supplementary Table S8.

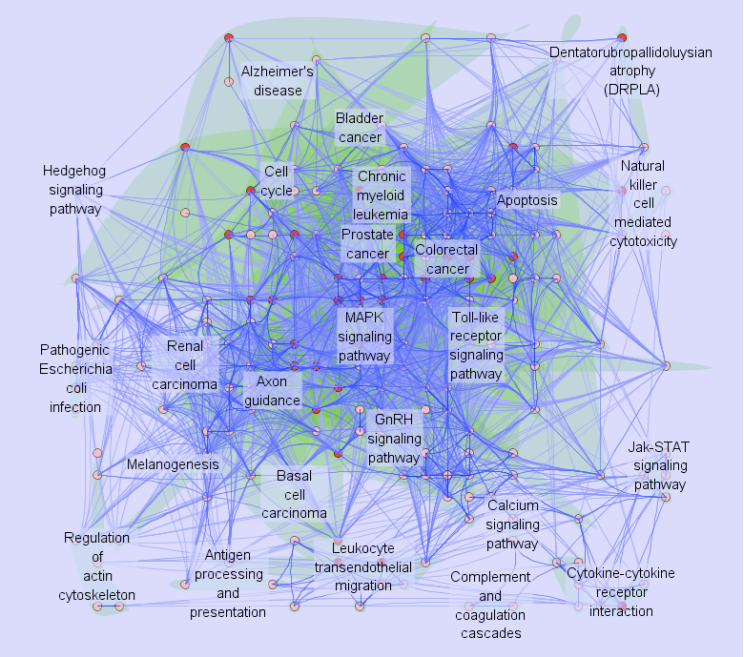


Figure 41. Extended network of the pathway crosstalks for HIV-induced neurodegeneration. Nodes in red color represent crosstalk proteins.

7.4. Discussion

Accumulating evidence suggests the striking similarities between the long-term microbial after-effects on memory loss or cognitive decline and characteristic symptoms of AD, which is in favor of the "infectious agent" hypothesis [244]. Indeed, the results from a

randomized clinical trial investigating the effects of antibiotics on patients with AD indicated that a combination of two common antibiotics could delay the cognitive decline [245].

In this study we applied our text mining technology to disease mechanism identification and pathway-pathway map reconstruction for virus-induced dementia with the focus on the HIV-associated neurodegeneration. The reason for choosing AIDS-dementia complex as the case study was two-fold: it is very well studied through extensive literature; and the manifestation of dementia in the course of disease is time-dependent.

Dementia is a syndrome which can be commonly seen in a number of different neurological disorders and clinical conditions ranging from infectious diseases such as AIDS to neurodegenerative diseases like Alzheimer's and Parkinson's. Neuronal cell death or neurodegeneration is widely observed in all these disorders but the exact factors or cofactors that initiate this process are not well understood. There have been several hypotheses regarding the mechanism of the neurodegeneration in dementia including the inflammatory hypothesis of dementia, the oxidative stress hypothesis, and the cell cycle hypothesis. However, we show here that all these hypotheses capture related parts of the same disease mechanism, although none has proved to uncover the causative factors upstream the disease mechanism. Moreover, we demonstrate that Alzheimer's and infection signaling pathways are connected in proximity to each other, which suggests that infection and dementia are probably associated at the molecular level.

Our knowledge-based approach suggests that all the above-mentioned hypotheses describe parts of a unified disease mechanism, which is initiated within the cytotoxic arm of the immune system. It is then speculated that microglia are activated by an unknown cause so that these resident innate immune cells in the brain in turn induce pro-inflammatory and oxidative stress pathways [246]. The latter pathway leads to DNA damage and consequently cell-cycle arrest through p53 and CDKNIA signaling pathways, which stimulates apoptosis and nerve cell death [247] (Figure 42).

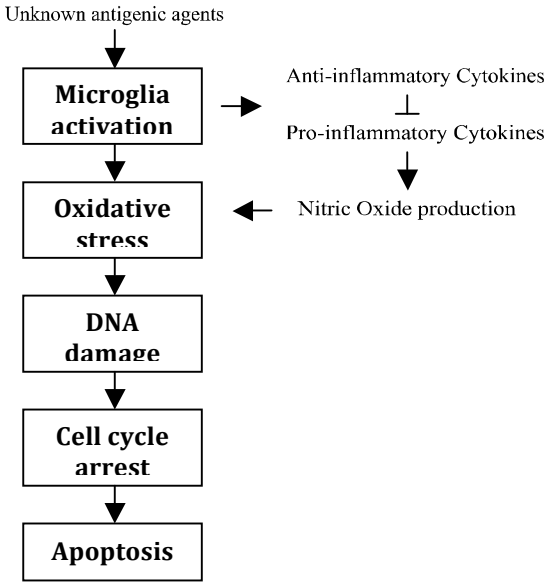


Figure 42. Schematic order of speculative events in HIV-induced neurodegeneration. This schema represents a hypothetical flow of conceptual causes in ADIS-dementia based on accumulated evidence in the literature.

It should be noted that this inference is highly speculative and is meant to elicit new mechanism-based hypotheses that fundamentally build on the infectious agent hypotheses of the AD etiology. Moreover, this inference does not explain the cause and initiation of AD but it merely builds on the molecular and pathway knowledge to guide how AD could potentially be initiated. With the availability of more data, it would be possible to test the proposed infection hypothesis and its link to AD through, e.g. systematic collection of external evidence (e.g. from GWAS studies) in the future.

As shown in Figure 41, the pathway interaction map is enriched by the immune response (cytokine-cytokine receptor interaction, complementary cascade, leukocyte migration, antigen processing and presentation, pathogenic E. coli infection pathway, Toll-like receptor signaling pathway, and Natural killer cell mediated cytotoxicity) and the DNA-damage apoptotic response (MAPK signaling pathway, cell cycle, different types of cancer pathways, and apoptosis). However, our approach identified another over-represented, yet hidden pathway in the literature known as Kynurenine pathway, which is not annotated in any of popular pathway databases. The kynurenine pathway is characterized for involvement in neuroprotection and regulation of the immune tolerance in human neurons via catabolism of L-tryptophan [248]. We found one study that shows decreased levels of kynurenic acid in serum and red blood cells of Alzheimer's disease patients [249]. Another earlier study concludes that activation of immune system increases the levels of kynurenines in the CSF of HIV-infected patients with neurological disorder [250]. Therefore, it will be encouraging to study the role and crosstalks of

the Kynurenine pathway in the context of neurodegenerative diseases.

One important point that must be taken into consideration is the phenomenon of time-dependency in all dementia-spectrum disorders. The cumulative evidence indicates that the neurotoxicity mediated by overactivated microglia is of a 'progressive' nature, so that such microglial cells remain overactive for years and continuous yet cumulative loss of neurons occurs over time [251],[252],[253],[254]. Similarly, virus-associated dementia appears in the late phases of infection after many years latency. We hypothesize that an unknown mechanism instigates microglia to secrete pro-inflammatory elements constitutively, which leads to disruption of the existing feedback loops between pro- and anti-inflammatory pathways.

Future investigations addressing this emerging hypothesis will be critical for the understanding of the causative agent or agents of dementia-spectrum diseases and the design of effective anti-dementia therapeutics.

7.5. Conclusion

Since our text mining and knowledge extraction system outputs a list of genes, which are ranked according to their mutual information content, the gene list analysis can be treated similar to the gene lists obtained from the expression studies. The gene list obtained from PubMed abstracts in this way represents two parts of the knowledge domain: the well-known intensively-researched part and the less-established recently-proposed part. Hereby, we used the former part as the backbone or layout of our framework to investigate the

relationships about the latter part. In this manner, disease-causing mechanisms can be studied in the context of pathway interaction networks, which may help generate novel hypotheses and identify those drug targets that are more likely to be involved in the causative effects of the disease.

While we focused on a case study of HIV-associated neurodegeneration in this work, our approach can be applied to virtually all kinds of disease. Building integrative models such as the one studied here helps researchers bring dispersed pieces of knowledge together and get a holistic picture of the problem under study. Clearly, elucidating the disease-causing mechanism for a certain phenotype or clinical outcome, which is manifested over a range of diseases that share similar clinical symptoms, could guide researchers to the causative factors underlying the core of diseases in question.

CHAPTER 8. Modeling of the Neuroendocrine Component associated to Alzheimer's Disease

As explained in Chapter 6, the exploratory data analysis indicated that hormone-signaling pathways could be involved in the pathology of AD. In this chapter, a first attempt towards modeling the neuroendocrine component of the AD mechanism has been described, which has been also published in the journal of Translational Medicine [255].

8.1. Background

The clinical hallmark of dementia-spectrum diseases including Alzheimer's disease (AD) and frontotemporal dementia is the gradual memory loss and impairment of other cognitive functions, which has been attributed to the aggregation of amyloid fibrils, a process known as amyloidogenesis [256],[257],[258]. However, recent findings indicate that many peptide and protein hormones are stored in secretory granules in the form of functional amyloid fibrils and such an amyloid-like structure is necessary for their natural functioning as hormones[259].

Moreover, observational studies on the beneficial effect of estrogen-based hormone therapy on cognitive impairment have also reported conflicting results [260]. Indeed, gender-specific risk profiles observed for dementia in elderly men and women have drawn

attention to the impact that sex hormones, as risk factors, may have on progression from mild cognitive impairment (MCI) to dementia [261],[262]. The higher risk of AD and dementia incidence in women has been linked to high serum levels of endogenous estrogen [263] and it has been shown to be influenced by hormone replacement therapy [264],[265],[266],[267],[268], although a better cognitive performance after current hormone therapy was dependent on the duration and type of treatment [269].

On the other hand, both basic and clinical research findings have consistently shown influence of a range of hormones on some cognitive functions in AD. For example, high levels of leptin in blood have been associated to a lower risk of AD [270] and leptin replacement therapy has been suggested as a novel therapeutic strategy for AD [271]. The loss of melatonin in cerebrospinal fluid has been observed in patients with dementia of Alzheimer's suggesting that it may play a role in the pathogenesis of AD [272],[273],[274],[275]. A low thyroid hormone level has been also associated with AD [276],[277]; the administration of thyroid hormone in AD model mice prevented cognitive deficit and improved the neurological function [278]. In Alzheimer's disease, a greater cognitive impairment has been found to be associated with lower CSF concentrations of corticotropin-releasing hormone [279],[280]. There is evidence that growth hormone (GH) declines with advancing age or in Alzheimer's disease [281],[282],[283] and that daily treatment of healthy older adults with GH improves the cognition independent of gender [284]. A recent study also shows that GH can boost memory retention in rats [285].

There are several lines of evidence that point to the role of insulin signaling in AD; e.g. insulin levels in the CSF of AD patients is lower than healthy controls [286],[287], insulin receptor signaling is compromised in AD neurons [288], and insulin resistance is associated with reductions in cerebral glucose metabolic rate, which is a risk factor for developing AD dementia [289]. Interestingly, epidemiological findings indicate that type II diabetes mellitus is linked to developing and exacerbating AD pathology [290],[291] so that Alzheimer's has been even proposed by some authors to be 'type III diabetes' [292],[293]. Similar neuroendocrine disturbances have been reported for Huntington's disease under which the thyrotropic, somatotropic and gonadotropic axes are altered [294].

All the above-mentioned evidence, including inconsistent results and disparate findings, suggests that there is a gap between the knowledge obtained from basic research and findings of clinical investigations on the association between hormones and cognition. Context-specific networks of molecular interactions provide a relevant framework for supporting translation of basic knowledge into clinically relevant information through integrative modeling of disease mechanism. Current Alzheimer's disease maps, including the recent AlzPathway model [295], lack the focused representation of hormone signaling pathways. Therefore, this work describes the first attempt to characterize the hormone/hormone-receptor interactions relevant to dementia disorders under a unified framework of the interconnected hormonal components.

8.2. Methods

Figure 43 summarizes the overall strategy used for this study. It demonstrates a top-down integrative (knowledge- and data-driven) approach to modeling the hormone protein interaction network.

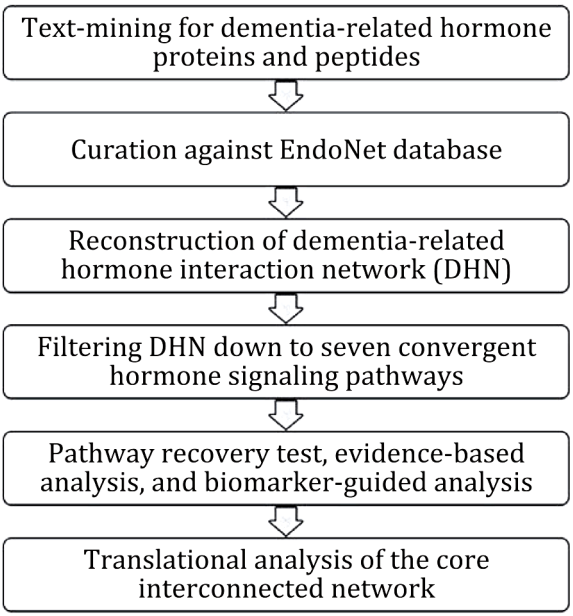


Figure 43. Schematic representation of the methodology used for construction and analysis of dementia-related hormonal network. Details on literature mining and model construction and analysis have been provided below.

8.2.1. Literature mining

Our retrieval system was composed of two software components: the dictionary-based text-mining tool, ProMiner, and the semantic search

engine, SCAIView. PubMed abstracts were searched for all instances of genes and proteins, which are mentioned in the context of 'dementia' as keyword (accessed as of 08.02.2011). The retrieved entities were manually checked for their true relevance to both hormones and dementia in the context of their abstracts.

8.2.2. Network reconstruction and annotation

The results from text-mining were cross-checked with the contents of the EndoNet database [296] and the confirmed entities were used as seed proteins in the BIANA tool for reconstruction of the dementia-related hormonal network (DHN) at level 2. The initial protein-protein interaction network was constructed around the seed set. In order to reduce the dimensionality and increase the confidence, interactions that are only supported by the yeast two-hybrid method were removed and those interactions that are independently confirmed by two or more experimental methods were maintained. The network was visualized and statistically analyzed in the Cytoscape and Gephi environments [297]. G-lay clustering algorithm was used for modularity analysis [298].

8.2.3. Pathways used for the recovery test

For the pathway recovery test, we obtained the following expert-curated hormonal pathways, used them as gold standard, and mapped them onto the network: growth hormone pathway [299], insulin signaling pathway [299], leptin signaling pathway [300], thyroid hormone signaling [301], regulation of the estrogen receptor pathway

[299], corticotropin-releasing hormone pathway [302], and Melatonin signaling pathway [303].

8.2.4. Statistical analysis

Gene set enrichment analysis was performed using the Molecular Signature Database (*MSigDB*). The DAVID functional annotation tool was used for annotation of differentially expressed genes in the network.

8.2.5. Translational validation

For establishing the clinical relevance of the core DHN model, knockout mouse phenotypes were retrieved from MGI database (http://www.informatics.jax.org/). For retrieval and extraction of putative biomarker information from the literature, the biomarker terminology developed and published by our team [204] was used. Pathway membership for each target was obtained from KEGG database (http://www.genome.jp/kegg/) and their association to disease was determined using genetic association database (geneticassociationdb.nih.gov). Information on brain tissue specificity of the targets was obtained from Tissue Distribution Database (genome.dkfz-heidelberg.de/menu/tissue_db/) but the higher resolution information at the cell type level was retrieved from the literaure using SCAIView search engine. DrugBank was searched for the proteins of the core DHN as targets of approved drugs (http://www.drugbank.ca/). Then PubMed was searched for supporting evidence of positive off-target effects of non-dementia

drugs that showed potential implication of those drug-targets in improvement of dementia.

8.3. Results

8.3.1. Enrichment of dementia-related proteins for hormone signaling activity

Mining the knowledge space of the literature for proteins that are shown to play a role in dementia resulted in a list of 1960 entities, which were ranked based on their mutual information (see Methods). Due to the fact that high-dimensional information returned by retrieval systems inherits noise, the next step was to observe whether signals of hormonal proteins could be detected in this large list of entities. The gene set enrichment analysis (GSEA) of these proteins revealed under-represented signatures of hormone activities in pathway analysis as well as implicit but statistically significant presence of hormone-related regulatory gene sets in GO biological process annotations. However, at the level of GO molecular function, these signatures showed significant over-representation for hormone activity, neuropeptide hormones and hormone signaling pathways (Table 32).

Table 32. Analysis of the retrieved data based on enrichment for gene ontology, category of molecular function. The top five annotations indicate a significant enrichment of the results for hormone activity.

Gene set name	Description of annotation	P-value
Amine receptor activity	GO:0008227: Combining with a biogenic amine to initiate a change in cell activity.	3.61E-05
Hormone activity	GO:0005179: The action characteristic of a hormone, any substance formed in very small amounts in one specialized organ or group of cells and carried (sometimes in the bloodstream) to another organ or group of cells in the same organism, upon which it has a specific regulatory action.	1.46E-04
Copper ion binding	GO:0005507: Interacting selectively with copper (Cu) ions.	3.40E-03
Neuropeptide hormone activity	GO:0005184: The action characteristic of a neuropeptide hormone, any peptide hormone that acts in the central nervous system. A neuropeptide is any of several types of molecules found in brain tissue, composed of short chains of amino acids.	6.55E-03
Serotonin receptor activity	GO:0004993: Combining with the biogenic amine serotonin, a neurotransmitter and hormone found in vertebrates, invertebrates and plants, to initiate a change in cell activity.	1.40E-02

The results of this observation led us to raise the hypothesis that an endocrine interaction network may exist that substantially contributes to the pathology of the dementia-spectrum diseases. To investigate this hypothesis, we used text-mining and knowledge discovery technologies to narrow down our search for retrieval and extraction of instances of hormone proteins and their receptors,

which are cited in the literature (Medline abstracts) in relation to dementia. The focused search resulted in retrieval of 1329 documents and 453 protein entities extracted from them. Finally, 89 hormone/hormone-receptor entities were confirmed to play a role as hormone/hormone-receptor after crosschecking the retrieved entities with the contents of the EndoNet database as gold standard (see Methods). We use this initial set of proteins (seed set) as *prior knowledge* to build upon our integrative model.

8.3.2. Dementia-related hormone network (DHN) and its biological relevance

The initial protein-protein interaction network comprises of 6966 nodes (proteins) and 85997 edges (interactions) but after filtering the number of edges in DHN decreased to 83998. 6515 nodes form a giant connected component and the rest of 451 nodes are singletons without any connection; thus, for simplicity, we only consider the giant component for further analyses. Statistical analysis of the giant component of DHN shows that its node degree distribution could be fitted in the power law of the form $y=1092.8x^{-1.17}$ with an acceptable goodness of fit (R-squared value=0.856, Correlation=0.996). This indicates that the network is of biological nature.

The network clustering coefficient of 0.315 and immense distribution of the clustering coefficients around the nodes with more than 100 neighbors is suggestive of a modular organization consisting of several interconnected functional modules. The modularity analysis of the network revealed four major modules whose functional annotation using GSEA supports the notion of modular organization

underlying the network (Figure 44): the largest module with 2037 nodes (Figure 44A) is significantly enriched for regulation of transcription, the second module with 1540 interconnected proteins (Figure 44B) is significantly involved in hormone and receptor signaling, the third module with 1420 proteins (Figure 44C) is significantly annotated for GPCR signaling, and finally the fourth module containing 1312 nodes (Figure 44D) is enriched for protein translation and induction of apoptosis. These findings are consistent with the fact that hormone peptides are major ligands for GPCRs and through cellular signaling cascades, they regulate the transcription of target genes in the nucleus.

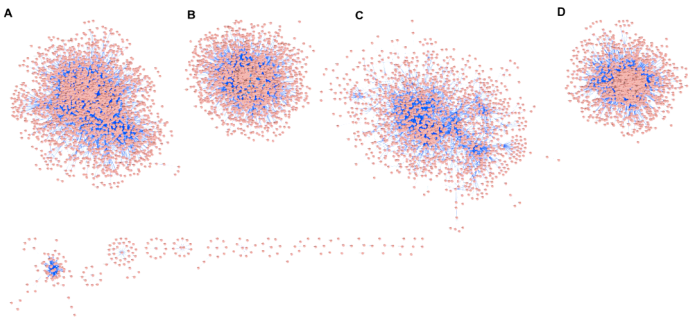


Figure 44. An overview of modules detected in DHN. 19 modules were detected in DHN out of which 4 modules represent ca. 97% of the network. (A) The largest module is enriched for regulation of transcription. (B) The second module with 1540 interconnected proteins representing hormone and hormone receptor signaling pathways. (C) The third module with 1420 proteins enriched for GPCR signaling. (D) The fourth module containing 1312 nodes is enriched for protein translation and induction of apoptosis.

8.3.3. Evaluation of DHN by pathway recovery

Both the biological relevance and the modularity were further evaluated by mapping the Alzheimer's disease pathway from the KEGG database as well as hormone signaling pathways from other resources (see Methods). Mapping the Alzheimer's disease pathway onto the network resulted in the recovery of all the proteins and their corresponding interactions in the pathway except for APH1A.

Regarding hormone signaling pathways, the number of proteins involved in the actual signaling and the number of mapped proteins for each signaling pathway is shown in Table 33. For two pathways with 100% node recovery, i.e. insulin signaling pathway and growth hormone pathway, manual extraction of edges (interactions) from BioCarta and mapping them onto the network yielded 76% edge recovery (16 out of 21) for the growth hormone pathway and 90% edge recovery (19 out of 21) for the insulin signaling pathway.

Table 33. List of dementia-related hormone signaling pathways that were recovered fully or partially in DHN

Hormone signaling pathway	No. of proteins in the original pathway	No. of proteins present in the hormone-dementia network
Estrogen receptor pathway	30	25 (83% recovery)
Insulin signaling pathway	21	21 (100% recovery)
Growth hormone pathway	27	27 (100% recovery)

Leptin signaling pathway	21	17 (80% recovery)
Thyroid signaling pathway	11	8 (72% recovery)
Melatonin signaling pathway	16	16 (100% recovery)
Corticotropin-releasing hormone signaling pathway	17	16 (94% recovery)

We also surveyed our network for the presence of hormone receptors by comparing them to known hormone receptors of genomic neuroendocrine hormones and were able to identify them for majority of these hormones.

8.3.4. Hormonal convergence in DHN

After the completion of this individual pathway recovery test, we aggregated all the elements of these seven pathways and mapped them onto the giant component of DHN. The aim was to detect the core of DHN where the majority of hormone crosstalks occur. A subnetwork of 73 nodes and 133 edges was formed, representing the converged hormonal pathway interactions. Interestingly, 62 of these hormone peptides (ca. 86%) are densely interconnected and form the core of DHN. Besides, their interactions appeared to occur in different regions of the normal brain after adding the context of brain region annotations to each edge using the brain interactome (Figure 45).

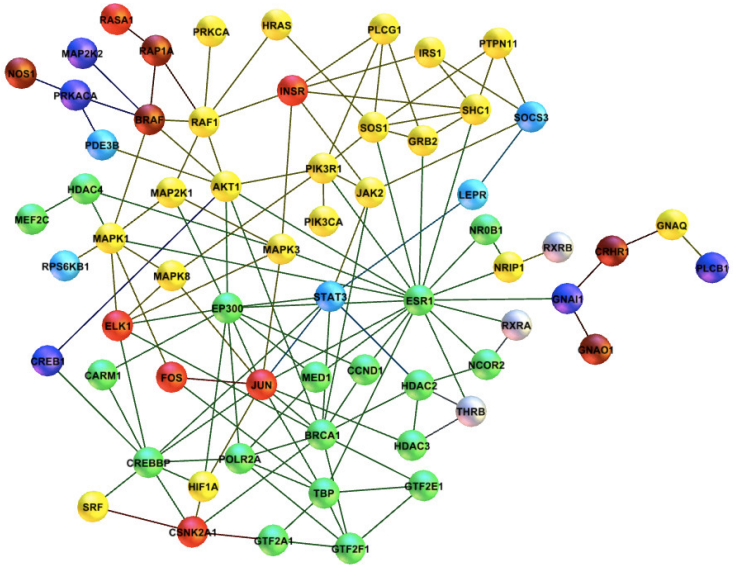


Figure 45. Elements of the seven hormonal signaling pathways form the core connected component of the brain interactome (normal state with all possible interactions). Pathway memberships are indicated by color codings; Green: estrogen signaling pathway; Red: insulin signaling pathway; Light blue: leptin signaling pathway; Dark blue: melatonin signaling pathway; Gray: thyroid signaling pathway; Brown: corticotropin-releasing hormone pathway. Yellow color indicates common membership to two or more pathways and also embeds the elements of the growth hormone signaling pathway.

Analysis of these annotations shows that the majority of the hormonal interactions occur in prefrontal cortex (ca. 93%), hypothalamus (ca. 92%) and cingulate cortex (ca. 90%), respectively. The finding that interactions of the converged network mostly occur in prefrontal cortex and cingulate cortex is consistent

with the neuroanatomical distribution of neurofibrillary tangles and plaques in the cerebral cortex of AD patients [304]. Moreover, the relevance of this finding to clinical attributes of the advanced AD pathology has been shown in several studies (**Prefrontal cortex:** [305],[306]; **Cingulate cortex:** [307]; **Hypothalamus:** [308],[309]. For example, it has been shown that prefrontal cortex, an important component for working memory, is the site of hormonal effects on cognition including estrogen [310], insulin [311], growth hormone [312], and thyroid hormone [313]. Thus, collective dysregulation of these pathways in prefrontal cortex of AD patients can lead to worsened memory impairment.

As the pathway-wise color codings in the converged hormonal network in Figure 45 indicates, a strong convergence and close interplay of hormone signals can be observed at the molecular level of the brain interactome. The yellow nodes show the common membership of proteins in two or more of these pathways and are significantly enriched for Neurotrophin/Trk signaling (GSEA p-value: $0e^{0}$, 14 genes in overlap), through which a variety of signaling cascades are connected and signals of neuronal development, survival as well as additional higher-order signals such as learning and memory are transmitted. The extended portion of estrogen signaling pathway in the core interactome is also noted.

8.3.5. Linking hormone-dementia hypothesis to mechanistic evidence

Apart from above *in silico* analyses, we provide more solid support for the hormone-dementia hypothesis from an Alzheimer's reference

expression data set [314], which has been processed and used for identification of a perturbed protein hub network in Alzheimer's disease by Liang et al. (2012) [315]. The Alzheimer's reference data set provides carefully phenotyped expression data set for six brain regions from late-onset AD patients (GSE5281) and lends support to the hypothesis that most of the differentially expressed genes in these six brain regions represent hub proteins in the hub network specific to Alzheimer's disease. We compared the core DHN with the Alzheimer's hub network derived from Alzheimer's reference expression data set and found 18 hormone signaling proteins in the core DHN that overlap with the hub genes differentially expressed in the hub network of Alzheimer's disease (Table 34). As table 34 indicates, all hormone signaling pathways are perturbed in different brain regions, with the largest overlap between insulin and growth hormone signaling pathways. Among these proteins, ESR1 and IRS1 exclusively represent two hormone signaling pathways, namely estrogen signaling pathway and insulin signaling pathway.

Table 34. Putative AD biomarkers in the core DHN supported by gene expression data and pathway membership. Red color indicates upregulation and green indicates downregulation of the corresponding genes, which are expressed in particular brain regions. Membership of those genes to their corresponding pathways are marked blue.

8.3.6. Translational validation of the core DHN

To our knowledge, except for hormone therapy with estrogen, there is no clinical trial describing the effect of other hormones on cognition improvement. Hence, in the absence of clinical trials, we propose a strategy for translational validation of the core DHN by showing the clinical relevance of the core DHN to dementia in the first step and then linking molecular signatures - through mouse model phenotypes - to their corresponding clinical manifestations.

The clinical relevance of the core DHN to dementia can be established through biomarker-guided analysis, in which information

of putative molecular indicators of dementia is retrieved and extracted from the literature and further become enriched with pathway membership, disease association and tissue/cell type specificity data (Table 35). Of the proteins in the core DHN, four were found in the literature to be reported as potential biomarkers that show measurable activity under Alzheimer's condition. These four proteins represent four different hormonal signaling pathways, namely growth hormone pathway (MAPK3), corticotropin-releasing hormone pathway (NOS1), melatonin signaling pathway (CREB1) and insulin pathway (JUN), whose measurable activities under AD condition suggest their mechanistic involvement in the pathology of AD dementia.

Table 35. Clinical relevance of the core DHN to dementia through biomarker-guided analysis

Target candidate	Pathway membership	Disease association	Biomarker type	Brain tissue specificity	Cell-type specificity
MAPK3	Alzheimer's disease, Prion disease, Type II diabetes mellitus, Insulin signaling pathway, Long-term potentiation	Autism	CSF increased levels AD (PMID: 22145083, 19625747), Phosphorylation (PMID: 19233276, 16920298, 17612901), Alterations in lymphoblasts of AD patients (PMID: 19158936)	Left ventricle, Right ventricle, Brain stem	Microglia, Astrocytes, Neurons

NOS1	Alzheimer's disease, Long-term depression, Calcium signaling pathway	Parkinson's disease, Alzheimer's disease, Diabetes Mellitus type II	Nitric oxide overproduction (PMID: 20804853), nNOS signaling initiated in interneurons (PMID: 16758165), Increased expression of nNOS isoforms in astrocytes (PMID: 12384247)	Substantia nigra, Forebrain, cerebral white matter, Limbic system	Astrocytes, Neurons
CREB1	Huntington's disease, Cholinergic synapse	Alzheimer's disease	Impaired CREB phosphorylation (PMID:22119240)	Substantia nigra, Brain stem, Sub-commissural organ, Brain ventricle, Cerebral gray matter, Cerebral white matter, Forebrain, limbic system	Hippocampal neurons, Dendate gyrus
JUN	GnRH signaling pathway, Neurotrophin signaling pathway, MAPK signaling pathway	Cognitive performance	Prolonged expression of c-Jun (PMID:8774439), Increased immuno-reactivity (PMID:8313943)	Sub-commissural organ, Brain ventricle, Cerebral gray and white matter	Neurons, Microglia, Substantia nigra

Next, we sought to investigate the translational value of DHN by linking hormone proteins in DHN to their corresponding knockout

mouse phenotypes. Table 36 summarizes 19 knockout mouse models representing 6 hormonal signaling pathways with phenotypes related to the nervous system. It also includes the ratio of knockout studies reporting an effect on the nervous system to studies reporting no effect on the nervous system.

Table 36. Knockout mouse phenotypes observed for several proteins in the core DHN model

Name	Mutation category	Observed effects on the nervous system	Ratio of KO studies with CNS phenotypes to studies without CNS phenotypes
Estrogen signaling			
Esr1^{tm1Ksk}	Targeted (knock-out)	abnormal pituitary gland physiology abnormal hypothalamus morphology abnormal innervation	4:8 (50%)
Ncor2^{tm1Kjep}	Targeted (knock-out)	abnormal cerebral cortex morphology abnormal neuron differentiation	1:0 (100%)
Hdac2$^{tm1.2Rdp}$	Targeted (knock-out)	abnormal hippocampus CA1 region morphology abnormal dentate gyrus morphology abnormal hippocampus pyramidal cell morphology enhanced long term potentiation	1:3 (34%)
Ccnd1^{tm1Wbg}	Targeted (knock-out)	absent Purkinje cell layer abnormal cerebellar granule layer small cerebellum	1:3 (34%)
Crebbp^{tm1Sis}	Targeted (knock-out)	abnormal forebrain morphology	3:4 (75%)
Insulin signaling			
Jun^{tm1Wag}	Targeted (knock-out)	abnormal forebrain morphology	1:2 (50%)
Hras1^{tm1Grnt}	Targeted	reduced long term	1:3 (34%)

	(knock-out)	potentiation	
Csnk2a1	Targeted (knock-out)	abnormal telencephalon development	1:2 (50%)
Mapk3^{tm1Gpg}	Targeted (knock-out)	reduced long term potentiation	2:2 (100%)
Leptin signaling pathway			
Lepr$^{tm1.2Chua}$	Targeted (knock-out)	abnormal inhibitory postsynaptic currents	1:4 (25%)
Stat3^{tm1Aki}	Targeted (knock-out)	abnormal motor neuron morphology abnormal neuron physiology	1:6 (17%)
Hif1a$^{tm1.1Stom}$	Targeted (knock-out)	abnormal cerebrum morphology abnormal cerebral cortex morphology loss of cortex neurons abnormal occipital lobe morphology abnormal temporal lobe morphology loss of hippocampal neurons	4:7 (57%)
Thyroid signaling pathway			
Rxrb^{tm1Rev}	Targeted (knock-out)	abnormal excitatory postsynaptic potential reduced long term potentiation absent long term depression	2:3 (67%)
Corticotropin-releasing pathway			
Gnaq^{tm1Soff}	Targeted (knock-out)	abnormal glutamate-mediated receptor currents absent long term depression	1:2 (50%)
Braf$^{tm1.1Sva}$	Targeted (knock-out)	increased neuron apoptosis abnormal innervation thin cerebral cortex	3:3 (100%)
Nos1^{tm1Plh}	Targeted (knock-out)	abnormal brain wave pattern abnormal long term potentiation	3:4 (75%)

		reduced long term potentiation absent long term depression decreased synaptic glutamate release abnormal peripheral nervous system regeneration	
Melatonin signaling pathway			
Gnai1^{tm1Drs}	Targeted (knock-out)	abnormal long term potentiation	1:1 (100%)
Plcb1^{tm1Hssh}	Targeted (knock-out)	loss of hippocampal neurons	1:1 (100%)
Creb1^{tm1Gsc}	Targeted (knock-out)	abnormal CNS synaptic transmission reduced long term potentiation	1:2 (50%)

To establish the bridge between the observed mouse phenotypes and the clinical disease manifestation in human, we propose the novel concept of "mechanism discovery through serendipitous off-target effects" based on the secondary positive effect of approved drugs that leads to unexpected and serendipitous clinical observations. Many approved drugs that are routinely used for treatment of human diseases lead to manifestation of so-called 'hidden phenotypes' due to binding to unknown targets [316]. The revelation of hidden phenotypes points to the fact that off-target effects sometimes result in positive effects through novel mechanisms of action. The most prominent example is the positive effect of Sildenafil on erectile dysfunction while the drug had been originally developed against angina.

Here we have collected a number of drugs with reported serendipitous effects on cognition that target several proteins in the core DHN (Table 37). Interestingly, all the off-targets of these drugs, when compared to the knockouts in Table 36, correspond to a nervous system phenotype in mice. Landscape illustration of the off-target effects in the core DHN model in Figure 46 shows that modulation of the estrogen signaling pathway by four drugs was more likely to lead to the serendipitous off-target effect on dementia and consequently, to improved cognitive functions observed in patients treated with these drugs.

Table 37. Drugs with serendipitous off-target effects on cognition and memory

Drug name	Main indication	Positive side effect on cognition and memory	Study subjects/ design	Supporting evidence (PMID)	Target protein in core DHN
Bexarotene (Targretin)	Skin cancer	Rapid reversal of cognition, social and olfactory deficits	Mouse model of AD	22323736	RXRA RXRB
Tamoxifen	Breast cancer	Higher level of independence in activities of daily life and decision making; relationship of tamoxifen with a lower prevalence of AD	Cross-sectional study of women receiving tamoxifen	11005221	ESR1
Raloxifene	Breast cancer	Reduced risk of cognitive impairment in	The Multiple Outcomes of Raloxifene	15800139	ESR1

		postmenopausal women	Evaluation randomized, placebo-controlled trial among postmenopausal women with osteoporosis		
Vorinostat	Cutaneous T cell lymphoma (skin cancer)	Complete restoration of contextual memory	Mutant APPswe/PS1dE9 mice	20010553	HDAC2 HDAC3
Lovastatin	Hyperlipid-emia	Reduction of Abeta formation and slowing the progression of AD	Double-blind randomized study on human subjects	11900994	HDAC2
Resverat-rol	Aging	Promoting clearance of Abeta peptides	Various cell lines	16162502	CSNK2A1
Sorafenib	Renal cell carcinoma	Reversal of memory impairment	Transgenic APPswe mouse model	20201822	BRAF
Naloxone	Opioid overdose	Improvement of learning and memory through enhancement of long-term potentiation	Aged rats with declined memory	14670637 15805661	CREB1

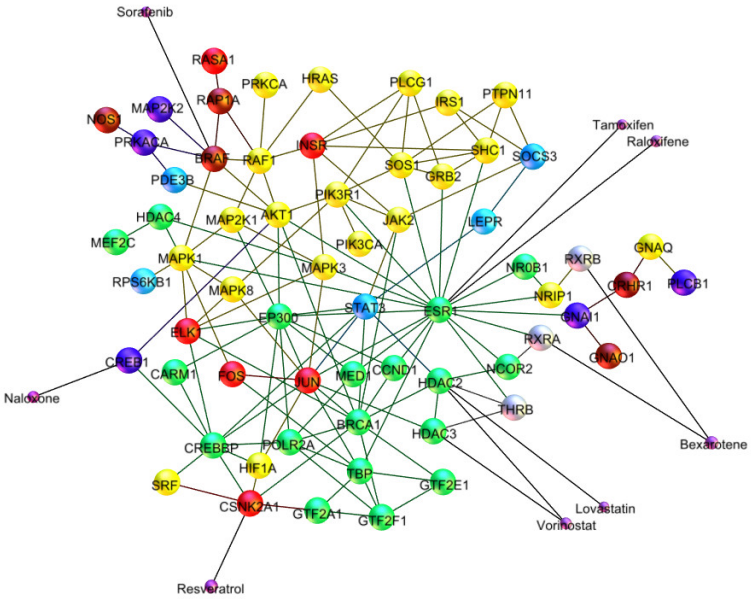


Figure 46. Schematic off-target landscape of 8 non-dementia drugs in the core DHN model. This enriched model is a proof-of-concept demonstrator for the mechanistic relevance of hormone signaling pathways to pathology of dementia.

To enlighten the usability of the DHN model, we performed a more general analysis by systematically searching for non-dementia drugs with targets of the core DHN model and retrieving published studies that support the positive, negative or neutral effect of those drugs on cognition or memory or learning. This analysis demonstrated that of 62 proteins in the core DHN model, 21 (ca. 33%) have been already targeted by at least one drug out of which 18 drugs targeting 13 proteins have shown positive effect, 3 drugs targeting 1 protein have

demonstrated negative effect, and 39 drugs targeting 18 proteins have not been investigated in relation to cognitive functions or have not been reported in the literature to have any observed effect on dementia. 21 proteins (ca. 33%) have been targeted by experimental compounds and 20 proteins (ca. 32%) have not been targeted by any drug or compound. These findings imply that hormone signaling pathways present a promising target space for drug discovery in the area of neurodegenerative brain disorders.

8.4. Discussion

Integrative modeling approaches provide a suitable medium for fusion of complementary data derived from literature and experiments. Given the fact that molecular mechanism of disease risk factors is often unclear and the exact mode of action of most approved drugs is unknown in most cases, such models can be used to interpret disease mechanism and to predict drug mode of action. In particular, an integrative model allows for inference of crosstalks among components of the system, guidance of analysis to the core pathological pathways, and generation of further hypotheses on how risk factors or disease-modifying treatments at molecular level lead to manifestation of positive or negative clinical effects. Accordingly, our in silico approach to modeling hormone signaling pathways that underlie dementia pathology provided several novel insights beyond what is already known about hormone signaling pathways in dementia, as follows.

The growing number of findings on the role of hormone signaling pathways in regulation of cognition and memory raises an immediate

question: how these bits and pieces of accumulating knowledge are being used to explain the contribution of hormones to improvement or exacerbation of dementia? The dementia-related hormonal network, presented in this paper, provided a first unified picture of the hormonal component underlying cognitive impairment. Convergence of genomic hormonal pathways in the DHN model uncovered tight molecular interconnections and crosstalks among hormone signaling pathways and regulatory pathways of neural growth, survival and differentiation. For instance, the observed convergence of estrogen and neurotrophin signaling pathways at the core of DHN has been shown to regulate an array of cytoskeletal and growth-associated genes in cerebral cortex, including tau microtubule associated protein, MAPT [317]. The implication of such hormone signals in the pathology of dementia is supported by the evidence that phosphorylation of MAPT, which leads to neurofibrillary tangle formation and ultimately neurodegeneration, is regulated by the signaling effects of insulin and estradiol [318],[319]. Similarly, the regulatory influence of thyroid hormone, melatonin, and corticotropin-releasing factors on hippocampal tau phosphorylation has been documented in the literature [320],[321],[322].

The DHN model could guide the mechanism discovery analysis to those signaling pathways that constitute the core pathological processes. The modularity detected in the network implies that hormone receptors and hormone signals in concert with transcription factors may play a significant part in the disease mechanism. The molecular interconnection of insulin pathway to dementia pathology

- revealed by the DHN model – may provide a mechanistic explanation for the previous epidemiological studies on the contribution of diabetes mellitus and insulin resistance as risk factors to exacerbation of dementia (e.g. [323],[324],[325]). For instance, very recently, a 9-year prospective study on 3069 elderly adults without dementia demonstrated that patients who suffered diabetes had significantly worse cognitive decline in comparison with those who did not have the disease, suggesting the contribution of diabetes mellitus severity to accelerated cognitive impairment [326]. An interesting observation in our model is the co-occurrence of diabetes-related proteins in the convergent core of DHN. Indeed, three members of this subnetwork (i.e. MAPK1, INSR, SOCS3) belong to the Type II diabetes mellitus pathway, which fall into the bigger insulin signaling pathway together with SHC1 and ELK1. As was shown by pathway recovery analysis, the insulin signaling pathway is present in the network with the highest number of nodes and edges amongst other signaling pathways. The presence of MAPK1/ELK1/CREBBP axis in the core subnetwork and its direct crosstalk to the insulin pathway is consistent with experimental observations that link insulin signaling and diabetes risk to the regulation of learning and formation of long-term memory [327],[328],[329].

We showed that the DHN model could have more valuable implications beyond a sole portrait of networked signaling pathways by enabling high-resolution analysis of core molecular events. This was achieved through enhancement of the DHN model with knockout phenotype data and drug-target information. Genetically

engineered mouse models play an instrumental role in studying disease mechanism and translating preclinical studies to the clinic [330]. Thus, the knockout phenotypes are good candidates for establishing the link between the molecular mechanism and the disease clinical manifestation. One clear observation from knockout phenotypes is the prominent involvement of all hormone signaling pathways in long term potentiation (LTP) beside other biological processes. It is well known that long term potentiation of synaptic transmission substantially contributes to memory formation [331] and that LTP inhibitors also block memory and learning [332]. Hence, it can be inferred from the model that probably perturbation in hormone signaling pathways may affect LTP adversely. Interestingly, knockout models of three putative biomarkers for dementia, namely MAPK3, NOS1 and CREB1, show reduced LTP, which supports the notion, that hormone signaling pathways are part of the dementia pathology. It should be noted that these proteins generally exert multiple functions in the biology of nervous system by participating in different signaling pathways and thus, the core DHN model describes their contribution to the hormone-mediated signaling in the context of dementia.

Since mouse knockout phenotypes alone might not be sufficient to concretely conclude about the translational value of our DHN model, introduction of the “serendipitous off-target effect” for linking DHN model to disease mechanism demonstrated to provide further validation for the DHN model. It was shown that inhibition of off-targets belonging to hormone signaling pathways could lead to improvement of memory and learning in human or animal models.

Therefore, the enhanced DHN model can be used to predict novel targets out of off-targtes or to identify disease-modifying targets and pathways that partially regulate the pathology of disease. For example, HDAC2 knockout mouse models show enhanced LTP, which may indicate HDAC2 might be a potential therapeutic target; on the other hand, the clinical evidence is provided by the off-target effect of Lovastatin on HDAC2: originally designed against hypercholesterolemia, Lovastatin was tested during a double-blind, randomized clinical study on human subjects for its effect on progression of Alzheimer's disease through reduction of amyloid-beta formation[333]. The study found that Lovastatin decreases the risk of AD progression. Such an inference exemplifies how the novel knowledge on the mechanism of drug effect on disease-related risk factors can be derived from the enhanced integrative model of DHN.

Although DHN provides a unified integrative map of possible hormone signaling mechanism in the context of dementia, it has its limitations. The inherent issue of network biology is that completeness of molecular network maps is limited to data availability and validity. The DHN model analyzed in this work may not cover all the hormone pathways involved in the pathogenesis of dementia but rather it focuses on the convergent hormone action by the most prominent ones. Furthermore, such models provide only a static picture and do not capture the dynamic behavior of the system. However, context-specific modeling, as the first step, makes it possible to simulate disease-specific perturbations after incorporation of quantitative data from high-throughput technologies. Such an integrative modeling approach may prove valuable for prediction of

potential biomarkers due to the fact that hormones are able to cross the blood-brain barrier by transmembrane diffusion or using transporters and their brain levels reflect blood levels [334]. We plan to keep the DHN model up to date – within the boundaries of available resources – by implementing an alert system that automatically collects new information published on the role of hormone signaling in dementia and enriches the model with the emerging knowledge. It is anticipated that, with the availability of more data, the resolution (i.e. specificity and sensitivity) of the model will increase so that new versions of the model will support translational scientists to make informed decisions.

8.5. Conclusions

The integrated hormone interaction model presented in this study can be beneficial in correlating the information of genes, proteins, signaling pathways and the clinical manifestation of dementia in the context of endocrine system. Such models have great potential to support the process of identifying new targets and novel biomarkers and bear the potential to help pharmaceutical industry to increase the efficiency of their pipeline.

CHAPTER 9. Modeling of the CNS Component and Temporal Progression of Alzheimer's Disease

Given the fact that efficient translation of valuable information embedded in brain scans into clinical application is of paramount scientific and public health importance, a strategy is proposed to bridge the current gap between imaging and molecular biology, particularly in AD. This work appeared in Nature Scientific Reports, November 2013 [335].

9.1. Introduction

Recent advancements in structural and functional neuroimaging techniques offer unprecedented opportunities to visualize brain structure and function, to non-invasively monitor the progression of a disease over time, or to track disease trajectories. Different types of imaging reveal different aspects of the brain complexity: Magnetic Resonance Imaging (MRI), Computerized Tomography (CT) and Diffusion Tensor Imaging (DTI) are designed to localize anatomical areas and structures (structural imaging techniques) whereas functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET) are used to capture neural activity at molecular level (functional imaging techniques) [336],[337]. Such imaging technologies have been used to identify structural and

functional changes associated with different stages of progressive neurodegenerative diseases such as Alzheimer's disease (AD) [338].

In AD patients, progressive loss of memory and cognitive abilities is attributed to the dysfunction and death of nerve cells in specific regions of the brain [339]. Imaging techniques have established the existence of such link between brain structural and functional changes by demonstrating the spatio-temporal patterns of cell death across affected brain regions. Differential patterns of brain atrophy observed in the brain of AD patients with the help of imaging techniques indicate that distribution of particular structural changes in specific regions of the AD brain may reflect the underlying pathology [340].

Based on the above-mentioned capabilities of imaging technologies, a steadily increasing number of imaging studies has been published on diagnosis and prognosis of AD but the reported applications are still limited to clinical monitoring of anatomical lesions or injuries of brain in the course of the disease. On the other hand, the assessment of persons with brain disorders and diagnostic decision-making process for such patients is still highly dependent on the skills of examiner and the patient's abilities, which shows the current limitations of brain imaging techniques for informing the diagnosis beyond the behavioral assessments [341]. Current automated methods for diagnosis of neuropsychiatric disorders make use of classification algorithms to classify the brain scans of participants based on measurements of local variation in the morphological features of the brain [342]. Accordingly, the diagnostic information derived from such imaging-based methods is often unspecific and the

knowledge behind the molecular mechanism underlying the measured imaging outcomes remains implicit or unknown. In fact, the challenge of linking clinical outcomes to their underlying molecular events has been long of interest to the scientific community and to the pharmaceutical industry as well. This is because it will help to obtain better understanding of the disease mechanism at molecular level, particularly for personalized medicine applications. The technique of molecular imaging using reporter molecules that provide information on particular molecular or cellular events has been around for some while but it is not in clinical use yet and its diagnostic as well as prognostic application will be limited to tracing of single cell or single cellular process [343].

To the best of our knowledge, there is no suggestion in the scientific or patent literature preceding the present study how to facilitate diagnosis and prognosis of brain diseases by translating information from a plurality of brain scan images to underlying region-specific disease pathways. Therefore, it would be desirable to provide a method that is able to not only diagnose brain images more accurately with higher specificity to the disease but also improve prognosis by prescribing efficient and personalized therapies based on medical history of individual patients. Such a method could be potentially upgraded to a clinical decision-support system that would adjoin neuroimaging softwares. It could also support target identification and biomarker discovery efforts as well.

The present work proposes a novel strategy using an integrative computational approach, which incorporates the information of imaging and potential protein biomarkers specific to disease into a

brain-specific protein interaction network. Enrichment analysis for known pathways further validated the method and unveiled the high impact of immune system on the pathology of AD.

9.2. Materials and methods

9.2.1. Information retrieval and extraction

With the help of state-of-the-art text mining and knowledge discovery tools, ProMiner and SCAIView, PubMed abstracts were searched using our dedicated biomarker terminology. The queries were formulated and executed over the complete set of PubMed abstracts on 14.05.2013 in SCAIView search engine (accessible through http://bishop.scai.fraunhofer.de/scaiview/). An example of query formulation is shown below:

(([MeSH Disease:"Alzheimer Disease"]) AND [BioMarker Ontology:"Diffusion tensor imaging"])

Similarly, with the help of SCAIView and the biomarker terminology, the following query was performed to obtain a list of potential AD biomarkers:

(([MeSH Disease:"Alzheimer Disease"]) AND [BioMarker Terminology Node:"Evidence Marker"])

The long list of retrieved potential AD biomarkers was filtered for expression evidence and was subjected to manual inspection of statements made in the paper abstracts.

Imaging abstracts were manually screened so that context (relevance to both AD and imaging biomarker) as well as content (information detailing the type of imaging biomarker and the affected brain

region) of retrieved documents was checked and the relevant information was extracted.

9.2.2. Reconstruction of the temporal lobe subnetwork

A brain-specific protein-protein interaction (PPI) network representing 15 brain regions was used (i.e. brain interactome). The PPI network was then filtered for the affected brain regions to obtain region-specific subnetwork for temporal lobe, based on the edge attributes using Cytoscape software.

9.2.3. Pathway enrichment analysis

Since the affected region in the early stage was the same in the advanced stage, we only analyzed one subnetwork representing temporal lobe in this study. The subnetwork and corresponding mapped molecular biomarkers were subjected to pathway enrichment analysis (GSEA) in MsigDB. In order to normalize the pathway matching process, GSEA was performed on both potential biomarkers and subnetwork proteins using separate pathway annotation sets, namely BioCarta, KEGG, and Reactome. In order to make the more specific enrichment results from the small number of potential biomarkers comparable to the less specific but high dimensional enrichment results from the large number of subnetwork proteins, we considered the top 20 enrichment results for subnetwork proteins so that the sensitivity of the pathway matching process is preserved.

9.3. Results

The core methodology for translation of imaging readouts to molecular pathway maps consists of two steps:

Step 1 intends to integrate the information of both imaging and molecular biomarkers into a brain-specific network model (so-called brain interactome), which represents experimentally confirmed protein interactions (i.e. network edges) in 15 anatomical regions of the human brain[9]. Since reports on brain imaging contain meta-information about clinical specifications of patient subjects such as stage of the disease and the affected region of the brain, the idea is to ultimately generate specific disease subnetworks out of the brain interactome that represent protein interactions in affected regions of the diseased brain. This is achieved by mapping information of imaging outcomes onto the region-specific brain interactome (Figure 47). For validation purposes, we have used imaging information reported in the literature on AD but the source of image-based diagnosis could in essence be extended to the clinician's medical report or any other diagnostic annotation attached to images.

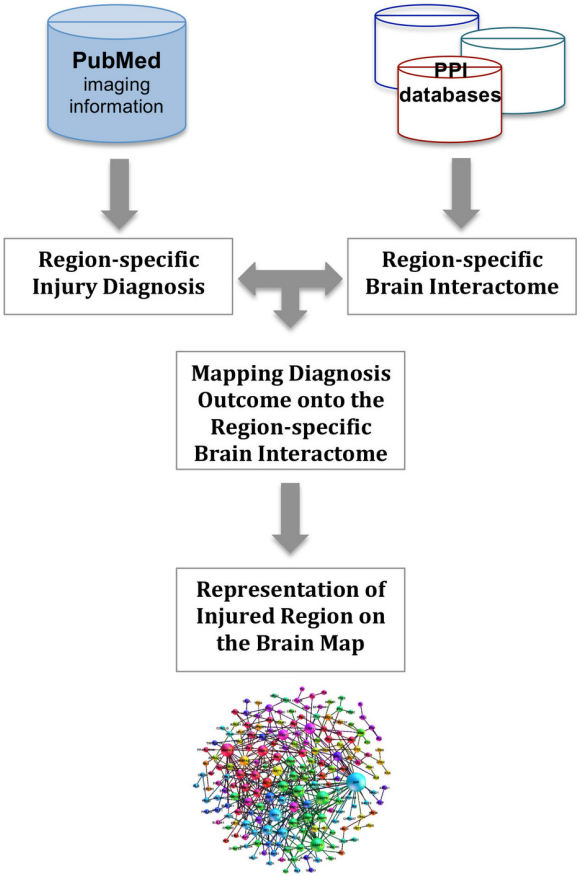


Figure 47. Generation of brain region-specific subnetwork models using imaging readouts. Region-specific diagnosis of brain injury can be matched with the region-specific subnetworks of the brain interactome to generate an interaction model specific to the injured region, in the first step.

Step 2 involves extraction of identified subnetworks from the brain interactome based on the affected regions that are diagnosed by imaging and consequently their analysis for underlying pathways.

The pathway analysis is performed on both the entire subnetworks and potential biomarkers mapped onto these subnetworks separately. Afterwards, pathways that are derived from subnetworks and pathways that are derived from mapped biomarkers are being matched so that potential molecular biomarkers act as "pins" on the disease map to guide the analysis to the core biological processes deemed to drive the pathology of the disease (Figure 48).

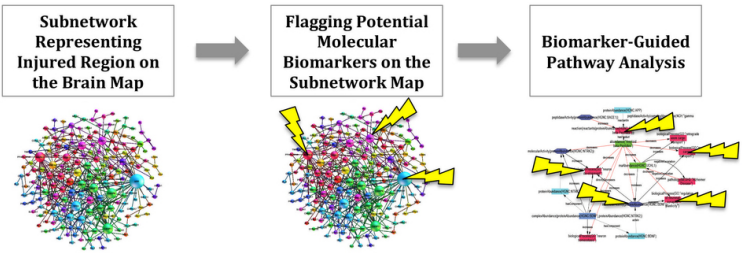


Figure 48. Enhancement of subnetwork models with the information of potential biomarkers. Region-specific subnetwork model generated based on the imaging diagnosis in the first step is subjected to biomarker-guided analysis by mapping potential biomarkers and pathway matching, as detailed below.

9.3.1. Method validation using imaging readouts of Alzheimer's patients

Our semantic information retrieval system, SCAIView, retrieved 5698 PubMed abstracts reporting clinical application of imaging techniques for diagnosis of AD (MRI: 3458, PET: 1989, DTI: 251) and containing information on both AD and affected brain regions (see Methods). After manual inspection of these abstracts and

information extraction, it turned out that the reported brain regions injured in AD could be assigned to three stages of the disease, namely early AD/mild cognitive impairment (MCI), moderate Alzheimer's, and advanced Alzheimer's. It was evident from the frequency of AD imaging publications that the majority of these studies reported the application of imaging methods to diagnosis of the early stage Alzheimer's disease, reflecting the high priority that is assigned to finding early diagnostics for AD. Moreover, these efforts were heavily biased toward application of MRI techniques amongst others.

Analysis of the diagnosed anatomical regions in AD brains indicated that both structural and functional techniques report limbic system and its anatomical components (i.e. temporal lobe, hippocampus, cingulate, thalamus and corpus callosum) as the affected regions across disease stages. Based on these readouts, a disease progression trend is observed so that it appears with progression of the disease from early to advanced stage, anatomical lesions extend from temporal lobe, entorhinal and precuneus cortices to prefrontal and cerebral cortices. These readouts suggest that temporal lobe lesion and dysfunction is persistent across disease stages. Therefore, we generated a temporal lobe subnetwork model out of the brain interactome and validated our proposed method on this model. The temporal lobe model is represented by a protein-protein interaction (PPI) subnetwork with 2323 nodes and 3587 edges.

9.3.2. Biomarker-pathway coupling for targeted enrichment analysis

In order to spot pathways causally involved in the progression of AD in the temporal lobe subnetwork model, we searched for molecular indicators (potential biomarker candidates) of AD in the literature, extracted a list of such hypothetical AD biomarkers (see Methods) and mapped them onto the temporal lobe subnetwork. Biomarkers are molecular alterations that can be measured in human tissue, cells or fluids and represent direct steps in the causal pathways of a disease. As a result, 144 potential biomarkers, including inflammatory and non-inflammatory biomarkers, were mapped onto the temporal lobe subnetwork. Since these potential biomarkers indicate measurable molecular activities under the AD condition, this strategy helps us guide our analysis to those pathways that are more likely to be involved in the disease mechanism at the molecular level. Moreover, it overcomes the challenge of dealing with the large number of significant pathways that often result from pathway enrichment analysis algorithms, particularly when a large number of proteins participating in network models are submitted for analysis.

To this end, we performed separate pathway enrichment analyses on both, the subnetwork proteins and the list of mapped biomarkers. Since the enrichment results from mapped biomarkers are used to guide the analysis on the subnetwork model, enrichment analysis on the biomarker list was performed using BioCarta, KEGG, and Reactome separately. The same analyses were performed on the subnetwork proteins. In this way, the same set of pathway annotations from each pathway database is used for pathway

comparison and the sensitivity of the pathway matching process between biomarker-derived pathways and subnetwork-derived pathways is maximized. The purpose is to find out which pathways are suggested by potential biomarkers to be perturbed in the subnetwork model and which pathways in the subnetwork model correctly represent the disease mechanism.

Such a biomarker-guided pathway analysis showed that imaging biomarkers point to the involvement of six pathways in progression of AD, namely HIV-NEF pathway, FAS signaling pathway, IL2RB pathway, keratinocyte pathway, MAPK signaling pathway and immune system signaling (Table 38).

Table 38. Matched results of pathway enrichment analysis on the temporal lobe subnetwork model

Source	Enriched pathways resulted from mapped biomarkers (FDR value)	Matched pathways in the subnetwork (FDR value)
BioCarta	HIV-I NEF pathway (0 e^{0}) FAS signaling pathway (2.41 e^{-14}) IL2RB pathway (1.33 e^{-13}) Keratinocyte pathway (4.94 e^{-11})	HIV-I NEF pathway (0 e^{0}) FAS signaling pathway (0 e^{0}) IL2RB pathway (0 e^{0}) Keratinocyte pathway (0 e^{0})
KEGG	MAPK signaling pathway (1.03 e^{-14})	MAPK signaling pathway (0 e^{0})
Reactome	Immune system (0 e^{0})	Immune system (0 e^{0})

Figure 49 illustrates HIV-NEF pathway spotted on the temporal lobe subnetwork model. In this model, there are two approved CNS drugs that target two proteins within the HIV-NEF pathway: Triflusal, which targets NFKB1 and is used for treatment of cerebral infarction and prevention of stroke; and Rasagiline, which targets BCL2 and is applied for treatment of idiopathic Parkinson's disease. Such a drug-

target-disease pathway landscape informs which therapies already target a disease pathway in a particular brain region. The specificity of the translated model increased even more when the expression information of mapped potential biomarkers (i.e. overexpressed or underexpressed) under AD conditions was also extracted from the literature and incorporated into the subnetwork model.

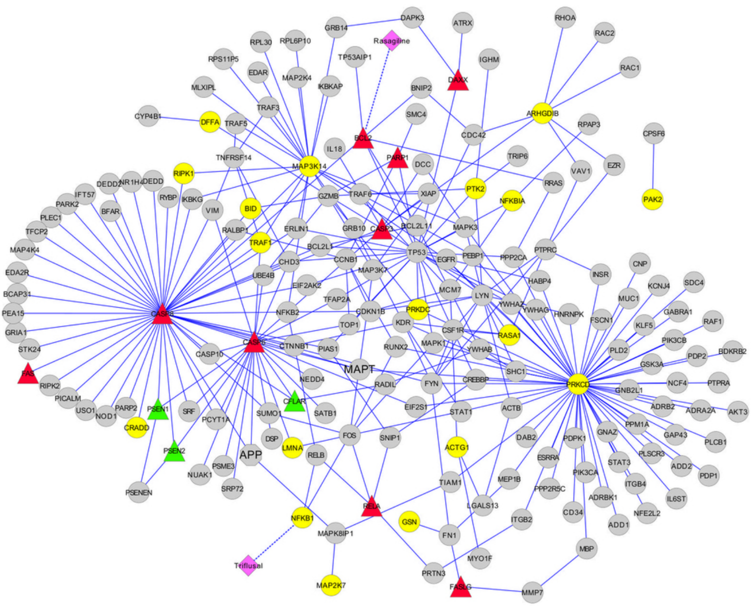


Figure 49. Representation of recovered HIV-NEF pathway and its first neighbor proteins in the temporal lobe network model. The model has been enhanced with drug-target and biomarker expression information. Circular nodes in yellow show membership to the HIV-NEF pathway; square nodes in pink are approved drugs targeting the recovered pathway; triangle nodes represent potential biomarkers color-coded for their expression levels in AD brain (red: over-expression; green: under-expression).

9.4. Discussion

In spite of invaluable contribution of neuroimaging to the understanding of disease progression, its outcome can not be directly used in the context of molecular systems analyses for translational purposes. The presented methodology is a novel approach to integrating brain imaging readouts into a network model of brain molecular interactions, which was validated using the accumulated knowledge on diagnostic neuroimaging of Alzheimer's disease in the literature. In this way, the pictorial information of brain scans that are not amenable to molecular analysis can be incorporated in a region-specific brain interaction network to analyze the resultant mechanistic models and to validate those models in the context of molecular pathways. Consequently, application of this approach to identifying drug targets can have important implications in CNS drug discovery by reducing the risk of drug failure in clinical trials, given the fact that it uses human imaging data instead of disease-mimicking animal data. Even at the level of sample data, a clear advantage of imaging biomarkers over gene expression signatures in neurodegenerative disease research is that they refer to in-vivo observations of regions and tissues in the diseased brain that are directly involved in the disease initiation and progression whereas gene expression signatures only provide a snapshot of perturbed genes, suffer from heterogeneity of cell types and are limited to post-mortem sampling.

Pathway enrichment analysis on the image-translated molecular map of temporal lobe revealed several pathways that were not previously appreciated to be causally involved in the pathogenesis of AD. The

advantage of using pathway enrichment analysis is that the collective effect of reported molecular biomarkers under the disease condition is taken into account in the context of disease pathways. When taken together, these pathways unveiled an important aspect of Alzheimer's pathology: immune system-driven apoptosis. Significant enrichment of temporal lobe subnetwork model and its mapped biomarkers for HIV infection pathway might indicate that signaling pathways to neuronal damage and apoptosis are elicited from the very early stage of AD, which persist over the period of advanced phase. The role of IL2RB and FAS signaling pathway in immune system-mediated apoptosis complements accumulated evidence that MAPK signaling pathways contribute to the pathogenesis of AD through regulation of neural apoptosis [344],[345],[346].

Enhancing these translated models with drug-target and biomarker information – as shown for HIV-NEF pathway model in Figure 49 – can provide added value to physicians and researchers in several ways:

- it may support more accurate diagnosis based on molecular etiology of the disease, particularly when measured biomarkers from patient are available and can be mapped onto the model to spot disease pathways. This has implications for the mechanistic diagnosis of diseases rather than conventional diagnosis solely based on often overlapping symptoms and signs.
- it may improve prognostic tasks using the drug-target information that is associated to disease pathways. Patient's therapeutic history can complement the model and support prognostic decision-making

through incorporation of individual risk factors such as susceptibility background (e.g. APOE genotype) or environmental risk factors (e.g. aging).

- it may guide target identification through prediction of drug mode of action in the context of affected tissue, disease stage and perturbed pathway. Information of approved drugs and their targets in disease pathways that are already incorporated in the model can support the concept of polypharmacology for discovery and development of next-generation multi-targeting drugs.

- it may be used for prediction of companion biomarkers that are mechanistically linked to disease etiology, on one hand, and to mode of action of approved or experimental drugs, on the other hand.

In summary, the novel integrative methodology presented here provides insight into the underlying molecular mechanisms of disease progression by linking the clinical readouts of imaging techniques to their corresponding molecular events, but this approach has limitations. The inherent issue of network biology is that completeness of molecular network maps is limited to data availability and validity. Therefore, the specificity and sensitivity of the translation process is a function of the completeness of the brain interactome. Another limiting factor is the low resolution of the protein-protein interaction maps in terms of representing other molecular species than proteins and also directionality of the interactions. However, these shortcomings can be overcome through replacement of PPI networks with causal computational models based on BEL. BEL-based mechanistic models not only represent all

molecular species such as ions or metabolites but also preserve the directionality of interactions.

This method is generic and can be applied to modeling other brain disorders. It can be foreseen that the extended algorithm of this methodology, when optimized and fully automated, has the potential to be used as a clinical decision-support tool for personalized diagnosis and prognosis of patients with brain disorders.

CHAPTER 10. Model-based Target Prediction for Alzheimer's Disease

10.1. Targetability versus druggability

In the book "*in silico* technologies in drug target identification and validation", Leon and Markel (2006) classify data-driven methods such a s expression profiling under the category of 'target identification' and assign text-mining and network/pathway analysis methods to the category of 'target validation' [347]. In this work, a hybrid approach to *in silico* identification and validation of drug targets in the context of AD has been introduced. The advantage of this hybrid approach, which combines data- and knowledge-driven strategies, is that it provides a unified framework for simultaneous identification and validation of potential target and biomarker candidates specific to the context of AD *in silico*.

In the opinion of this author, it is important to distinguish between "targetability" and "druggability" features. While most of the studies have been focused on the druggability properties of the protein targets, less attention has been paid to the targetability properties of protein targets. This notion is supported by the fact that the primary target for 7% of approved drugs is not known and mode of action for 18% of approved drugs is not defined [348],[349]. Druggability is defined as "the ability of a target to be modulated by potent, small drug-like molecules" [350], mostly reflects the structured-based

physicochemical properties of the target in the binding site [351], and is used in the target validation phase [352]. In contrast, there is no clear definition for targetability in the literature so far. Targetability can be defined as 'the ability of a target to modify the path of disease or modulate disease-related phenotypes'. It is often used in the target identification phase [353].

The concept of targetability is being transformed from a 'target-based' paradigm into 'pathway-based' paradigm, where network subgraphs and pathways emerge as targetable entities (e.g. [354],[355]). Rising attrition rates of new compounds in the past decade, which was highest (62%) during phase II [356], indicates the lack of efficacy and reflects the low predictive capacity of target-based discoveries. Advantages of the pathway-based approach over the target-based approach are manifold [357]:

a) the hypothesis behind a target's mechanism of action in the context of the disease can be disproved (i.e. what if manipulation of target X fails to modify the disease process Y);
b) the functional output of the target pathway can be predicted and linked to clinically relevant outcomes; and
c) positive therapeutic off-target effects of approved or pipeline drugs can be predicted.

Several studies have previously shown the effectiveness of using pathways as therapeutic targets in neurobiology. For example, measurement of hippocampal neurogenesis pathway by high-throughput screening for approved drugs on mouse models showed that cholesterol lowering drugs can predict the stimulatory effect of these drugs on the adult neurogenesis pathway in animal models

[358]. In another study by the same group, lipopolysaccharide-induced microglia proliferation pathway in the rat brain was subjected to HTS analysis and drugs were found that ameliorated clinical symptoms in the mouse models of Parkinson's disease [359]. In a primary study based on pathway analysis, Cramer et al. (2012) found that FDA-approved anti-cancer drug, bexarotene, could rapidly reverse cognitive and olfactory deficits in Alzheimer's mouse models [360]. Inspired by these results, Li and Lu (2013) have proposed a pathway-based drug repositioning prediction method, which considers all possible causal chains connecting drugs to diseases through molecular pathways and estimates transition likelihood of each causal chain in the network [361].

10.2. Application of the hybrid approach to target identification and validation for AD

As mentioned in Chapter 6, the triad of Brain, Immune, and Hormone signaling components create an inter-linked and complex system, which is an example of integrated and interactive system where nervous, immune/inflammatory, and endocrine components preserve complex crosstalks in order to guarantee the maintenance of the system's homeostasis. Due to such a high level of integrity, dysregulation or malfunctioning of one component directly affects the neighbouring components. Separate hypotheses on the involvement of each component in the pathology of AD have been proposed but they all describe the impairment in components of the same system. Therefore, the collective results from modeling

components of the triad system should be taken into account when the identification of a targetable pathway is desired.

The proposed hybrid approach provides a systems view on the involvement of the triad system in the pathology of AD. Interplay between the immune, nervous and endocrine components has attracted attention of researchers for many years and it is now well established that this triad interaction is mediated by cytokines, which regulate important processes such as cell proliferation and hormone secretion [362]. Accordingly, immunotransmitters transmit information from the immune component to the CNS component and interfere with the neuroendocrine component. This triad interaction may explain, at least in part, impairments in immunity against infections, reproductive functions, and brain atrophy, which occur in the pathophysiology of AD.

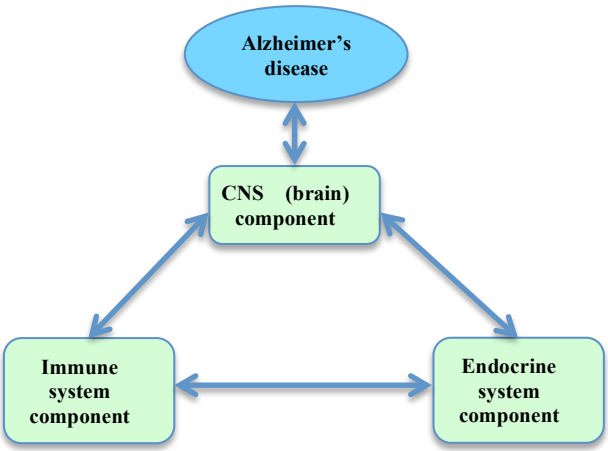


Figure 50. The triad system hypothesized to be involved in the pathology of AD and repreentation of their interactions.

When analyzing the pathway enrichment for literature-derived proteins reported to be involved in dementia (Chapter 8), it came to the attention that the Biocarta's cytokines and inflammatory response pathway is the most significant overrepresented pathway. On the other hand, modeling results from both AIDS-dementia complex and AD progression studies (Chapters 7 and 9, respectively) are indicative of the prominent role of the immune response signaling and cellular regulatory signaling pathways in the core pathology of AD. More weight should be given to the results of biomarker-guided pathway analysis in the AD staging study because *in-vivo* imaging biomarkers in combination with *in-vitro* measured molecular biomarkers reflect a more realistic picture of the disease development in AD patients.

Comparison of the enriched pathways that resulted from modeling of each component implies that MAPK signaling pathway is common among all the three components. In the hormone-dementia study above, MAPK3, which regulates the growth hormone pathway, represents the MAPK signaling pathway. It has been shown that MAPK signaling pathway is pathologically involved in various human diseases including AD, PD and ALS [363]. This is because mitogen-activated protein kinases mediate intracellular signaling and regulate a diverse set of cellular activities. Besides, it has been shown that extra-cellular amyloid-beta activates MAPK signaling cascade via nicotinic receptors [364]. In AD, MAPK pathway enforces abnormal re-entry into the cell cycle, which in turn activates amyloid production pathways; more specifically, MAPK3 is activated by oxidative stress, which is a hallmark of

neurodegeneration [365]. MAPK signaling pathway has been suggested as a potential therapeutic target for neurodegenerative diseases because protein kinases are attractive modulators of brain inflammation and gliosis and MAPK inhibitors can be easily administered due to good bioavailability [366]. In comparison to the field of oncology where half of the drugs in the pipeline are kinase inhibitors, CNS disease indications appear to be lagging behind for kinase-targeted drugs. For example, Imatinib (Gleevec), which is used for treatment of multiple cancers, was first reported for its efficacy in 1995-6 and was ultimately approved and introduced to the market in 2001 whereas no kinase inhibitor exists to date for CNS diseases. In fact, the launch of Imatinib demonstrated the success of applying pathway approaches in target validation. However, there are several kinase inhibitors that are either at the preclinical or clinical phase of investigation for treatment of CNS disorders [367].

Given the above results, however, an ideal therapeutic strategy seems to require modulating all the disease pathways underlying the triad system. So far, the mainstream of investigation for therapeutic interventions in AD has been focused on targeting amyloid production, accumulation, clearance, or toxicity associated with amyloid-beta plaques. Due to disappointments with these approaches, as discussed earlier, alternative approaches are under investigation, which include targeting oxidative stress, inflammation, and glutamate-mediated neurotoxicity; in addition, neurorestoration via neurotrophin pathway and hormone therapies are emerging approaches to modifying the path of AD [368]. Unfortunately, these

entire alternative approaches have been founded on the same old hypothesis of amyloidosis and each one of them represents a piece of the same puzzle. Acknowledging the complexity of AD, it becomes apparent that targeting single pathways through these approaches may not result in effective treatments for AD. Hence, it is proposed here that the concept of polypharmacology can be extended from the level of "multi-protein" targeting to the level of "multi-pathway" targeting.

CHAPTER 11. Conclusion and Outlook

Deciphering disease mechanisms is an important mission in systems biology and drug discovery. When dealing with multi-factorial progressive diseases such as neurodegeneration, it is of utmost importance to firstly get a profound understanding and knowledge on molecular mechanisms because this will lend a more support to predictive models for target identification and validation.

One of the biggest hurdles to finding novel biomarkers and therapeutics for brain diseases is the lack of deep understanding of the "etiological mechanism". As an example, the current classification of diseases is solely based on anatomical, symptomatic and epidemiological criteria and does not take the etiological mechanism into account. At the same time, neurological disorders may share common disease mechanisms despite heterogeneous clinical symptoms. Thus, re-defining diseases based on their underlying molecular and environmental causes rather than on physical signs and symptoms is crucial for development of effective therapeutic strategies. In addition, the elucidation of disease mechanism may support discovery of new disease-specific biomarkers.

Despite decades of research on various aspects of dementia, the mechanisms underlying neurodegeneration are yet far from being well understood. The fact is that years of pure data-driven approaches to delineating the complexity of disease mechanism in

dementia merely based on GWAS or gene expression results has not yielded a practical and impactful outcome. Polymorphism studies may be underpowered to detect rare variants, gene expression studies are highly sensitive to genetic, environmental, demographic and technical factors, and proteomic results – particularly in the biomarker discovery for dementia – are difficult to reproduce [369]. A general problem with most of data-driven methods is that they are devoid of proper disease-specific biological context; i.e. they ignore cell-type and tissue specificity, disease staging, and patient-specific risk factors at the time of sampling. The argument is that data-driven approaches alone simply are not amenable to understanding the complexity underlying NDDs, in general, and AD, in particular. Given the increasing amount of information density in the published knowledge space of literature, the task of knowledge-driven approaches is to bring context to multi-layered data-driven methods based on prior biological knowledge. This prior knowledge spans from published pathways and co-expression information to clinical parameters including age, comorbidities, and disease progression, to name a few.

As shown throughout this thesis, a knowledge-based integrative approach to modeling neurodegenerative disease mechanism in AD was designed and utilized successfully to link molecular disease states to their corresponding clinical readouts. A substantial amount of time and effort was dedicated to groundbreaking work and development of semantic frameworks specific to dementia and AD, which did not exist before. Various ontologies and controlled vocabularies were generated, evaluated, and applied to support the

knowledge-driven approach to modeling AD dementia mechanism. The complementary nature of knowledge-driven and data-driven approaches allowed for combining experimental data with prior knowledge under a single modeling framework, which ultimately resulted in enhanced sematic resolution at biological scales, improved functional sensitivity at the molecular level, and increased disease specificity at clinical level.

The results of disease modeling approach applied to AD in this work clearly show a deviation from the conventional amyloid-centered results. It appears that amyloid plaque formation and inflammation are secondary effects of an earlier damage to the neurons, which can be seen as a consequence of dysregulation in the immune system. Thus, the next wave of neurodegenerative research should be heavily directed towards investigation of the role of the immune system in susceptibility to and initiation of the disease. It is predictable that the next generation of computational disease models integrate medical histories and health outcomes of patients, for example in the form of co-morbidity analysis, so that deeper insights into the early events in the pathogenesis of NDDs is gained and preventive measures to modify the path of disease progression can be taken.

The capacity of *in silico* disease models to be augmented with annotation of semantic and mechanistic information such as anatomic sites, knock-out phenotypes, pathways, drug targets and/or potential biomarkers provides a flexible framework for development of algorithms that support decision making in drug development, clinical diagnosis and prognosis of diseases. The proposed algorithm for translation of brain scans to their underlying region-specific

molecular maps is an example of such augmentation efforts, which uses the novel "biomarker-guided analysis" method and thereby opens a new opportunity to computationally validate the *in silico* models of the disease. In should be noted that the disease models presented in this work are correlation, unsigned network models, which have been built using protein-protein interaction data. An emerging alternative in computational disease model building integrates literature-derived 'cause and effect' relationships into a data-driven platform and produces casual network models. Moreover, BEL-based disease models go beyond modeling of 'single-entity interactions' by including more biological entities as well as biological processes. An interesting feature of BEL models is the ability to identify the information flow between up-stream and down-stream of cellular processes [370]. Given this capability, it would be interesting to devise a strategy for reasoning over the BEL models aiming at simulating effects of hypothetical ligands or perturbations at the up-stream points and predicting outcomes at the down-stream endpoints.

By now, the concept of polypharmacology has been restricted to targeting multiple proteins in a single unique pathway. However, as discussed earlier in this chapter, the paradigm shift from single-target to pathway-target strategy demands for extension of the polypharmacology concept to the pathway level. The rationale supporting this proposition is that NDDs manifest heterogeneous clinical symptoms in various pathological locations and, therefore, targeting multiple disease pathways at the same time is more likely to exert a greater modifying effect on the mechanism of disease. As

shown in the present study, pathway-based disease modeling can be used for building the bridge from drug discovery to clinical research by linking molecular specifications of the disease mechanism to corresponding clinical manifestations. There are both advantages and limitations associated to this approach. Advantages include offering an alternative successful route to the current conventional target identification and validation in the context of disease pathway models, placing single genes and proteins in the context of pathway signatures for development of responder-specific therapeutics, and increasing the druggable space. Nonetheless, target relationships between pathways due to cross-talk events in the same cell or tissue, and different species-specific signaling mechanisms between distinct pathway models are among limiting factors. Moreover, in the absence of data and with limited "direct" AD knowledge, this sort of "loose association mining" is imperative. However, it can be foreseen that integrative disease modeling methods will be improved with advancement of “concrete association mining” algorithms as more data and information becomes available. In addition, it is anticipated that these methods will be used in the future to link the mode of action for drug candidates to predictable side effects or to build connectivity between in-vivo and in-vitro biomarkers.

REFERENCES

[1] Tascioglu, A. O., & Tascioglu, A. B. (2005). Ventricular anatomy: illustrations and concepts from antiquity to Renaissance. Neuroanatomy, 4, 57-63.

[2] Lagerlund, H. (Ed.). (2011). Encyclopedia of Medieval Philosophy. Springer Science.

[3] Silver RA, Kanichay RT (2008) Refreshing connections. Science 320:183-184.

[4] Stam, C. J. (2010). Characterization of anatomical and functional connectivity in the brain: a complex networks perspective. International Journal of Psychophysiology, 77(3), 186-194.

[5] Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186-198.

[6] Ikezu, T., Gendelman, H. E., & Przedborski, S. (Eds.). (2008). Neuroimmune pharmacology. New York: Springer.

[7] Turkington, C., & Harris, J. (2009). The encyclopedia of the brain and brain disorders. Infobase Publishing.

[8] Pritchard, C., Mayers, A., & Baldwin, D. (2013). Changing patterns of neurological mortality in the 10 major developed countries–1979–2010. Public health.

[9] Young, A. B. (2009). Four decades of neurodegenerative disease research: how far we have come!. The Journal of Neuroscience, 29(41), 12722-12728.

[10] Becker, R. E., & Greig, N. H. (2013). Fire in the ashes: Can failed Alzheimer's disease drugs succeed with second chances? Alzheimer's & Dementia, 9(1), 50-57.

[11] The Lancet. Why are drug trials in Alzheimer's disease failing?. Lancet 2010; 376: 658.

[12] Selkoe, D. J. (2011). Resolving controversies on the path to Alzheimer's therapeutics. Nature medicine, 17(9), 1060-1065.

[13] Herrup, K., Carrillo, M. C., Schenk, D., Cacace, A., DeSanti, S., Fremeau, R., ... & Budd, S. (2013). Beyond amyloid: Getting real about nonamyloid targets in Alzheimer's disease. Alzheimer's & Dementia, 9(4), 452-458.

[14] Gomeni, R. (2014). Use of predictive models in CNS diseases. Current Opinion in Pharmacology, 14, 23-29.

[15] Przedborski, S., Vila, M., & Jackson-Lewis, V. (2003). Series Introduction: Neurodegeneration: What is it and where are we?. Journal of Clinical Investigation, 111(1), 3-10.

[16] Dunnett SB, Björklund A. (1999) Prospects for new restorative and neuroprotective treatments in Parkinson's disease. Nature, 399:A32-A39.

[17] Singh-Manoux, A., & Kivimäki, M. (2010). The importance of cognitive ageing for understanding dementia. Age, 32(4), 509-512.

[18] Williams, A. (2002). Defining neurodegenerative diseases: Disorders will be named after responsible rogue proteins and their solutions. BMJ: British Medical Journal, 324(7352), 1465.

[19] Committee on A Framework for Developing a New Taxonomy of Disease: What would a knowledge network and new taxonomy look like? In Towards precision medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. The National Academies Press: Washington DC; 2011:35-50.

[20] http://www.scai.fraunhofer.de/geschaeftsfelder/bioinformatik-neu/projekte/ aetionomy.html

[21] Govoni, S., Bolis, L., & Trabucchi, M. (Eds.). (1999). Dementias: biological bases and clinical approach to treatment. Springer.

[22] Agronin, M. E. (2007). Alzheimer disease and other dementias: a practical guide. Wolters Kluwer Health.

[23] Kola, I., & Bell, J. (2011). A call to reform the taxonomy of human disease. Nature Reviews Drug Discovery 10(9), 641-642.

[24] World Alzheimer's Report 2009. London, Alzheimer's Disease International, 2009.

[25] Ritchie, K., & Lovestone, S. (2002). The dementias. The Lancet, 360(9347), 1759-1766.

[26] Mayeux, R., & Stern, Y. (2012). Epidemiology of Alzheimer disease. Cold Spring Harbor Perspectives in Medicine, 2(8).

[27] Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G., ... & Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science, 261(5123), 921-923.

[28] Brown, R. C., Lockwood, A. H., & Sonawane, B. R. (2005). Neurodegenerative diseases: an overview of environmental risk factors. Environmental health perspectives, 113(9), 1250.

[29] Sims-Robinson C, Kim B, Rosko A, Feldman EL: How does diabetes accelerate Alzheimer disease pathology? Nat. Rev. Neurol. 2010, 6:551-559.

[30] McGeer PL, McGeer E, Rogers J, Sibley J: Anti-inflammatory drugs and Alzheimer disease. Lancet 1990, 335:1037.

[31] Duthie, A., Chew, D., & Soiza, R. L. (2011). Non-psychiatric comorbidity associated with Alzheimer's disease. QJM, 104(11), 913-920.

[32] Grossman, H., Bergmann, C., & Parker, S. (2006). A Brief Review. The Mount Sinai journal of medicine, 73(7), 985.

[33] Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., ... & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. The Lancet Neurology, 6(8), 734-746.

[34] Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., ... & Scheltens, P.

(2010). Revising the definition of Alzheimer's disease: a new lexicon. The Lancet Neurology, 9(11), 1118-1127.

[35] Whitwell, J. L., Josephs, K. A., Murray, M. E., Kantarci, K., Przybelski, S. A., Weigand, S. D., ... & Jack, C. R. (2008). MRI correlates of neurofibrillary tangle pathology at autopsy A voxel-based morphometry study. Neurology, 71(10), 743-749.

[36] Jack Jr, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., ... & Trojanowski, J. Q. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology, 12(2), 207-216.

[37] Reisberg, B. (1987). Functional assessment staging (FAST). Psychopharmacology bulletin, 24(4), 653-659.

[38] Auer, S., & Reisberg, B. (1997). The GDS/FAST staging system. International Psychogeriatrics, 9(1), 167-172.

[39] Reisberg, B., Prichep, L., Mosconi, L., John, E. R., Glodzik-Sobanska, L., Boksay, I., ... & de Leon, M. J. (2008). The pre–mild cognitive impairment, subjective cognitive impairment stage of Alzheimer's disease. Alzheimer's & Dementia, 4(1), S98-S108.

[40] Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. Acta neuropathologica, 82(4), 239-259.

[41] Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., & Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta neuropathologica, 112(4), 389-404.

[42] Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., ... & Phelps, C. H. (2011). 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. Alzheimer's & Dementia, 7(3), 280-292.

[43] Hedden, T., & Gabrieli, J. D. (2004). Insights into the ageing mind: a view from cognitive neuroscience. Nature reviews neuroscience, 5(2), 87-96.

[44] Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. Neuron, 35(4), 625-641.

[45] Schulte, E., Schumacher, U., & Rude, J. (2007). Thieme atlas of anatomy: head and neuroanatomy (Vol. 3). Thieme.

[46] Barnes, C. (2011). Secrets of Aging-What does a normally aging brain look like? Are diseases of aging such as Alzheimer's inevitable?. Scientist, 25(9), 30.

[47] Bartsch, T., & Butler, C. (2013). Transient amnesic syndromes. Nature Reviews Neurology, 9(2), 86-97.

[48] Irish M, Addis DR, Hodges JR, Piguet O (2012) Considering the role of semantic memory in episodic future thinking: evidence from semantic dementia. Brain …

[49] Glisky EL (2007) Changes in cognitive function in human aging. In: Riddle DR, ed. Brain aging: models, methods, and mechanisms. Chap 1. Boca Raton, FL: CRC Press.

[50] Solari, S. V. H., & Stoner, R. (2011). Cognitive consilience: primate non-primary neuroanatomical circuits underlying cognition. Frontiers in neuroanatomy, 5.

[51] Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186-198.

[52] Zhou, J., Greicius, M. D., Gennatas, E. D., Growdon, M. E., Jang, J. Y., Rabinovici, G. D., ... & Seeley, W. W. (2010). Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. Brain, 133(5), 1352-1367.

[53] William Rebeck, G., Reiter, J. S., Strickland, D. K., & Hyman, B. T. (1993). Apolipoprotein E in sporadic

Alzheimer's disease: allelic variation and receptor interactions. Neuron, 11(4), 575-580.

[54] Deary, I. J., Whiteman, M. C., Pattie, A., Starr, J. M., Hayward, C., Wright, A. F., ... & Whalley, L. J. (2002). Ageing: Cognitive change and the APOE ϵ 4 allele. Nature, 418(6901), 932-932.

[55] Verghese, P. B., Castellano, J. M., & Holtzman, D. M. (2011). Apolipoprotein E in Alzheimer's disease and other neurological disorders. The Lancet Neurology, 10(3), 241-252.

[56] Alonso, A. D. C., Grundke-Iqbal, I., & Iqbal, K. (1996). Alzheimer's disease hyperphosphorylated tau sequesters normal tau into tangles of filaments and disassembles microtubules. Nature medicine, 2(7), 783-787.

[57] Alonso, A. D. C., Zaidi, T., Novak, M., Grundke-Iqbal, I., & Iqbal, K. (2001). Hyperphosphorylation induces self-assembly of τ into tangles of paired helical filaments/straight filaments. Proceedings of the National Academy of Sciences, 98(12), 6923-6928.

[58] Markesbery, W. R. (1997). Oxidative stress hypothesis in Alzheimer's disease. Free Radical Biology and Medicine, 23(1), 134-147.

[59] Manolopoulos, K. N., Klotz, L. O., Korsten, P., Bornstein, S. R., & Barthel, A. (2010). Linking Alzheimer's disease to insulin resistance: the FoxO response to oxidative stress. Molecular psychiatry, 15(11), 1046-1052.

[60] McShea, A., Wahl, A. F., & Smith, M. A. (1999). Re-entry into the cell cycle: a mechanism for neurodegeneration in Alzheimer disease. Medical hypotheses, 52(6), 525-527.

[61] Yang, Y., Mufson, E. J., & Herrup, K. (2003). Neuronal cell death is preceded by cell cycle events at all stages of Alzheimer's disease. The Journal of neuroscience, 23(7), 2557-2563.

[62] McGeer, P. L., & McGeer, E. G. (1995). The inflammatory response system of brain: implications for therapy of

Alzheimer and other neurodegenerative diseases. Brain Research Reviews, 21(2), 195-218.

[63] Wyss-Coray, T. (2006). Inflammation in Alzheimer disease: driving force, bystander or beneficial response?. Nature medicine, 12(9), 1005-1015.

[64] McGeer, P. L., & McGeer, E. G. (2013). The amyloid cascade-inflammatory hypothesis of Alzheimer disease: implications for therapy. Acta neuropathologica, 126(4), 479-497.

[65] Hardy, J., & Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science, 297(5580), 353-356.

[66] Tanzi, R. E., & Bertram, L. (2005). Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. Cell, 120(4), 545-555.

[67] Kung, H. F. (2012). The β-amyloid hypothesis in Alzheimer's disease: seeing is believing. ACS Medicinal Chemistry Letters, 3(4), 265-267.

[68] Dong, S., Duan, Y., Hu, Y., & Zhao, Z. (2012). Advances in the pathogenesis of Alzheimer's disease: a re-evaluation of amyloid cascade hypothesis. Translational neurodegeneration, 1(1), 18.

[69] Blennow, K. *et al*. (2010) Cerebrospinal fluid and plasma biomarkers in Alzheimer disease *Nat. Rev. Neurol.* doi:10.1038/nrneurol.2010.4

[70] Haass, C., & Selkoe, D. J. (2007). Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. Nature reviews Molecular cell biology, 8(2), 101-112.

[71] Andrew, F. T., & Ottavio, A. (2012). Is the Amyloid Hypothesis of Alzheimer's disease therapeutically relevant?. Biochemical Journal, 446(2), 165-177.

[72] Chételat, G. (2013). Alzheimer disease: Aβ-independent processes—rethinking preclinical AD. Nature Reviews Neurology.

[73] Armstrong, R. A. (2011). The pathogenesis of Alzheimer's disease: a reevaluation of the "amyloid cascade hypothesis". International journal of Alzheimer's disease, 2011.

[74] Dong, S., Duan, Y., Hu, Y., & Zhao, Z. (2012). Advances in the pathogenesis of Alzheimer's disease: a re-evaluation of amyloid cascade hypothesis. Translational neurodegeneration, 1(1), 18.

[75] Caselli, R. J., & Reiman, E. M. (2013). Characterizing the Preclinical Stages of Alzheimer's Disease and the Prospect of Presymptomatic Intervention. Journal of Alzheimer's Disease, 33, S405-S416.

[76] Bateman, R. J., Xiong, C., Benzinger, T. L., Fagan, A. M., Goate, A., Fox, N. C., ... & Morris, J. C. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. New England Journal of Medicine, 367(9), 795-804.

[77] Buchhave, P., Minthon, L., Zetterberg, H., Wallin, A. K., Blennow, K., & Hansson, O. (2012). Cerebrospinal Fluid Levels of {beta}-Amyloid 1-42, but Not of Tau, Are Fully Changed Already 5 to 10 Years Before the Onset of Alzheimer Dementia. Archives of general psychiatry, 69(1), 98.

[78] Weiner, M. W. (2013). Dementia in 2012: Further insights into Alzheimer disease pathogenesis. Nature Reviews Neurology, 9(2), 65-66.

[79] Frank-Cannon, T. C., Alto, L. T., McAlpine, F. E., & Tansey, M. G. (2009). Does neuroinflammation fan the flame in neurodegenerative diseases. Mol Neurodegener, 4(47), 1-13.

[80] McGeer, P. L., Schulzer, M., & McGeer, E. G. (1996). Arthritis and anti-inflammatory agents as possible protective factors for Alzheimer's disease A review of 17 epidemiologic studies. Neurology, 47(2), 425-432.

[81] McGeer EG, McGeer PL: Inflammatory processes in Alzheimer's disease. Prog Neuropsychopharmacol Biol Psychiatry 2003, 27(5):741-9.

[82] Glass, C. K., Saijo, K., Winner, B., Marchetto, M. C., & Gage, F. H. (2010). Mechanisms underlying inflammation in neurodegeneration. Cell, 140(6), 918-934.

[83] Holmes, C., Cunningham, C., Zotova, E., Woolford, J., Dean, C., Kerr, S., ... & Perry, V. H. (2009). Systemic inflammation and disease progression in Alzheimer disease. Neurology, 73(10), 768-774.

[84] vom Berg, J., Prokop, S., Miller, K. R., Obst, J., Kälin, R. E., Lopategui-Cabezas, I., ... & Heppner, F. L. (2012). Inhibition of IL-12/IL-23 signaling reduces Alzheimer's disease-like pathology and cognitive decline. Nature medicine.

[85] Heneka, M. T., & O'Banion, M. K. (2007). Inflammatory processes in Alzheimer's disease. Journal of neuroimmunology, 184(1), 69-91.

[86] MacDonald AB: Borrelia in the brains of patients dying with dementia. JAMA 1986, 256:2195-2196.

[87] Mawanda, F., & Wallace, R. (2013). Can infections cause Alzheimer's disease?. Epidemiologic reviews, 35(1), 161-180.

[88] Honjo, K., van Reekum, R., & Verhoeff, N. P. (2009). Alzheimer's disease and infection: do infectious agents contribute to progression of Alzheimer's disease?. Alzheimer's & Dementia, 5(4), 348-360.

[89] Nussbaum, J. M., Schilling, S., Cynis, H., Silva, A., Swanson, E., Wangsanut, T., ... & Bloom, G. S. (2012). Prion-like behaviour and tau-dependent cytotoxicity of pyroglutamylated amyloid-[bgr]. Nature, 485(7400), 651-655.

[90] Morales, R., Duran-Aniotz, C., Castilla, J., Estrada, L. D., & Soto, C. (2011). De novo induction of amyloid-β deposition in vivo. Molecular psychiatry.

[91] Perry, V. H., Cunningham, C., & Holmes, C. (2007). Systemic infections and inflammation affect chronic neurodegeneration. Nature Reviews Immunology, 7(2), 161-167.

[92] Mayeux, R., Stern, Y., & Spanton, S. (1985). Heterogeneity in dementia of the Alzheimer type Evidence of subgroups. Neurology, 35(4), 453-453.

[93] Cummings, J. L. (2000). Cognitive and behavioral heterogeneity in Alzheimer's disease: seeking the neurobiological basis. Neurobiology of aging, 21(6), 845-861.

[94] de Waal, H., Stam, C. J., Blankenstein, M. A., Pijnenburg, Y. A., Scheltens, P., & van der Flier, W. M. (2011). EEG abnormalities in early and late onset Alzheimer's disease: understanding heterogeneity. Journal of Neurology, Neurosurgery & Psychiatry, 82(1), 67-71.

[95] Janocko, N. J., Brodersen, K. A., Soto-Ortolaza, A. I., Ross, O. A., Liesinger, A. M., Duara, R., ... & Murray, M. E. (2012). Neuropathologically defined subtypes of Alzheimer's disease differ significantly from neurofibrillary tangle-predominant dementia. Acta Neuropathologica, 124(5), 681-692.

[96] Lam, B., Masellis, M., Freedman, M., Stuss, D. T., & Black, S. E. (2013). Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. Alzheimer's research & therapy, 5(1), 1-14.

[97] Armstrong, R. A., Nochlin, D. and Bird, T. D. (2000), Neuropathological heterogeneity in Alzheimer's disease: A study of 80 cases using principal components analysis. Neuropathology, 20: 31–37.

[98] Salomone, S., Caraci, F., Leggio, G. M., Fedotova, J., & Drago, F. (2012). New pharmacological strategies for treatment of Alzheimer's disease: focus on disease modifying drugs. British journal of clinical pharmacology, 73(4), 504-517.

[99] Mount, C., & Downton, C. (2006). Alzheimer disease: progress or profit? Nature medicine, 12(7), 780-784.

[100] Cummings, J. L., Miller, B. L., Christensen, D. D., & Cherry, D. (2008). Creativity and dementia: emerging

diagnostic and treatment methods for Alzheimer's disease. *CNS Spectr*, *13*(2 Suppl 2), 1-20.

[101] Mangialasche, F., Solomon, A., Winblad, B., Mecocci, P., & Kivipelto, M. (2010). Alzheimer's disease: clinical trials and drug development. The Lancet Neurology, 9(7), 702-716.

[102] Godfrey-Smith, P. (2006). The strategy of model-based science. Biology and Philosophy, 21(5), 725-740.

[103] Haefner, J. W. (2005). Modeling biological systems: principles and applications. Springer.

[104] Barillot, Emmanuel, Calzone, Laurence, Hupe, Philippe, ... & Andrei Yu Zinovyev. (2012). Computational systems biology of cancer (Vol. 47). CRC Press.

[105] Liu, E. T. (2005). Systems biology, integrative biology, predictive biology. Cell, 121(4), 505-506.

[106] Kitano, H. (2002). Systems biology: a brief overview. Science, 295(5560), 1662-1664.

[107] Dhar, P. K., Zhu, H., & Mishra, S. K. (2004). Computational approach to systems biology: from fraction to integration and beyond. NanoBioscience, IEEE Transactions on, 3(3), 144-152.

[108] Barabási, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 5(2), 101-113.

[109] Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998)

[110] Estrada, E. (Ed.). (2010). Network Science: Complexity in Nature and Technology. Springer.

[111] Snoep, J.L. (2005) The Silicon Cell initiative: working towards a detailed kinetic description at the cellular level. Curr. Opin. Biotechnol. 16, 336–343.

[112] Reed, J. L., Senger, R. S., Antoniewicz, M. R., & Young, J. D. (2011). Computational approaches in metabolic engineering. BioMed Research International, 2010.

[113] Bruggeman, F. J., & Westerhoff, H. V. (2007). The nature of systems biology. TRENDS in Microbiology, 15(1), 45-50.

[114] Small, D. H., San Mok, S., & Bornstein, J. C. (2001). Alzheimer's disease and Aβ toxicity: from top to bottom. Nature Reviews Neuroscience, 2(8), 595-598.

[115] Mizuno, S., Iijima, R., Ogishima, S., Kikuchi, M., Matsuoka, Y., Ghosh, S., ... & Tanaka, H. (2012). AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. BMC Systems Biology, 6(1), 52.

[116] Casadesus, G., Arendash, G., Laferla, F., & McDonald, M. (2011). Animal models of Alzheimer's disease. International journal of Alzheimer's disease, 2010.

[117] Zhu, X., Gerstein, M., & Snyder, M. (2007). Getting connected: analysis and principles of biological networks. Genes & development, 21(9), 1010-1024.

[118] Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nature Reviews Genetics, 12(1), 56-68.

[119] Lima-Mendez G, van Helden J: The powerful law of the power law and other myths in network biology. Mol Biosyst 2009, 5(12):1482-1493.

[120] Samal A, Martin OC: Randomizing genome-scale metabolic networks. PLoS ONE 2011, 6(7):e22295.

[121] Ingram PJ, Stumpf MP, Stark J: Network motifs: Structure does not determine function. BMC Genomics 2006, 7:108.

[122] Axelsen, J. B., Bernhardsson, S., & Sneppen, K. (2008). One hub-one process: a tool based view on regulatory network topology. BMC systems biology, 2(1), 25.

[123] Han, J. D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., ... & Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature, 430(6995), 88-93.

[124] Li, X., Wu, M., Kwoh, C. K., & Ng, S. K. (2010). Computational approaches for detecting protein complexes from protein interaction networks: a survey. BMC genomics, 11(Suppl 1), S3.

[125] Vidal, M., Cusick, M. E., & Barabasi, A. L. (2011). Interactome networks and human disease. Cell, 144(6), 986-998.

[126] Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. Science, 297(5586), 1551-1555.

[127] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102(43), 15545-15550.

[128] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists Nucl. Acids Res. (2009) 37 (1): 1-13.

[129] Lage, K., Karlberg, E. O., Størling, Z. M., Olason, P. I., Pedersen, A. G., Rigina, O., ... & Brunak, S. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. Nature biotechnology, 25(3), 309-316.

[130] Bauer-Mehren, A., Bundschus, M., Rautschka, M., Mayer, M. A., Sanz, F., & Furlong, L. I. (2011). Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. PloS one, 6(6), e20284.

[131] Fujita, K. A., Ostaszewski, M., Matsuoka, Y., Ghosh, S., Glaab, E., Trefois, C., ... & Balling, R. (2013). Integrating pathways of Parkinson's disease in a molecular interaction map. Molecular neurobiology, 1-15.

[132] Toedter G, Li K, Sague S, Ma K, Marano C, Macoritto M, Park J, Deehan R, Matthews A, Wu GD, Lewis JD, Arijs I, Rutgeerts P, Baribaud F: Genes associated with intestinal permeability in ulcerative colitis: Changes in expression

following infliximab therapy. Inflammatory bowel diseases 2012, 18:1399-1410.

[133] Westra, J. W., Schlage, W. K., Frushour, B. P., Gebel, S., Catlett, N. L., Han, W., ... & Hoeng, J. (2011). Construction of a computable cell proliferation network focused on non-diseased lung cells. BMC systems biology, 5(1), 105.

[134] Ideker, T., & Krogan, N. J. (2012). Differential network biology. Molecular systems biology, 8(1).

[135] Taylor, I. W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., ... & Wrana, J. L. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature biotechnology, 27(2), 199-204.

[136] Martin, F., Thomson, T. M., Sewer, A., Drubin, D. A., Mathis, C., Weisensee, D., ... & Peitsch, M. C. (2012). Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks. BMC systems biology, 6(1), 54.

[137] Kola, I., & Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates?. Nature reviews Drug discovery, 3(8), 711-716.

[138] Knight, Z. A., Lin, H., & Shokat, K. M. (2010). Targeting the cancer kinome through polypharmacology. Nature Reviews Cancer, 10(2), 130-137.

[139] Shen, Y. et al. Blueprint for antimicrobial hit discovery targeting metabolic networks. Proc. Natl Acad. Sci. USA 107, 1082–1087 (2010).

[140] Bolognesi, M. L., Cavalli, A., Valgimigli, L., Bartolini, M., Rosini, M., Andrisano, V., ... & Melchiorre, C. (2007). Multi-target-directed drug design strategy: from a dual binding site acetylcholinesterase inhibitor to a trifunctional compound against Alzheimer's disease. Journal of medicinal chemistry, 50(26), 6446-6449.

[141] Nacher, J. C., & Schwartz, J. M. (2012). Modularity in protein complex and drug interactions reveals new polypharmacological properties. PloS one, 7(1), e30028.

[142] Chen, Y., Zhu, J., Lum, P. Y., Yang, X., Pinto, S., MacNeil, D. J., ... & Schadt, E. E. (2008). Variations in DNA elucidate molecular networks that cause disease. Nature, 452(7186), 429-435.

[143] Arrell, D. K., & Terzic, A. (2010). Network systems biology for drug discovery. Clinical Pharmacology & Therapeutics, 88(1), 120-125.

[144] Csermely, P., Korcsmáros, T., Kiss, H. J., London, G., & Nussinov, R. (2013). Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. Pharmacology & therapeutics 138(3): 333-408.

[145] McDermott, J. E., Diamond, D. L., Corley, C., Rasmussen, A. L., Katze, M. G., & Waters, K. M. (2012). Topological analysis of protein co-abundance networks identifies novel host targets important for HCV infection and pathogenesis. BMC Systems Biology, 6(1), 28.

[146] Liu, H., Xiao, F., Serebriiskii, I. G., O'Brien, S. W., Maglaty, M. A., Astsaturov, I., ... & Connolly, D. C. (2013). Network analysis identifies an HSP90-central hub susceptible in ovarian cancer. Clinical Cancer Research, 19(18), 5053-5067.

[147] Engin, H. B., Keskin, O., Nussinov, R., & Gursoy, A. (2012). A strategy based on protein–protein interface motifs may help in identifying drug off-targets. Journal of chemical information and modeling, 52(8), 2273-2286.

[148] McDermott, J. E., Wang, J., Mitchell, H., Webb-Robertson, B. J., Hafen, R., Ramey, J., & Rodland, K. D. (2013). Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. Expert opinion on medical diagnostics, 7(1), 37-51.

[149] Azuaje, F. J., Dewey, F. E., Brutsaert, D. L., Devaux, Y., Ashley, E. A., & Wagner, D. R. (2012). Systems-based approaches to cardiovascular biomarker discovery. Circulation: Cardiovascular Genetics, 5(3), 360-367.

[150] Chuang HY, Lee E, Liu YT, Lee D, Ideker T: Network-based classification of breast cancer metastasis. Mol Syst Biol 2007, 3:140.

[151] Liu, X., Liu, Z. P., Zhao, X. M., & Chen, L. (2012). Identifying disease genes and module biomarkers by differential interactions. Journal of the American Medical Informatics Association, 19(2), 241-248.

[152] Wang K, Lee I, Carlson G, Hood L, Galas D: Systems biology and the discovery of diagnostic biomarkers. Disease markers 2010, 28:199-207.

[153] DUDLEY, JOEL T., and Atul J. Butte. "Identification of discriminating biomarkers for human disease using integrative network biology." Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing. NIH Public Access, 2009.

[154] Dunnett SB, Björklund A: Prospects for new restorative and neuroprotective treatments in Parkinson's disease. Nature 1999, 399:A32-A39.

[155] Loscalzo J, Kohane I, Barabasi AL: Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. Mol Syst Biol 2007, 3:124.

[156] Committee on A Framework for Developing a New Taxonomy of Disease: What would a knowledge network and new taxonomy look like? In Towards precision medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. The National Academies Press: Washington DC; 2011:35-50.

[157] Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. Nature chemical biology, 4(11), 682-690.

[158] Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I., Hufeisen, S. J., ... & Roth, B. L. (2009). Predicting new molecular targets for known drugs. Nature, 462(7270), 175-181.

[159] Takarabe, M., Kotera, M., Nishimura, Y., Goto, S., & Yamanishi, Y. (2012). Drug target prediction using adverse event report systems: a pharmacogenomic approach. Bioinformatics, 28(18), i611-i618.

[160] Hu, G., & Agarwal, P. (2009). Human disease-drug network based on genomic expression profiles. PLoS ONE, 4, e6536.

[161] Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J., & Bork, P. (2008). Drug target identification using side-effect similarity. Science, 321(5886), 263-266.

[162] Mizutani, S., Pauwels, E., Stoven, V., Goto, S., & Yamanishi, Y. (2012). Relating drug–protein interaction network with drug side effects. Bioinformatics, 28(18), i522-i528.

[163] Sams-Dodd, F. (2005). Target-based drug discovery: is something wrong?. Drug discovery today, 10(2), 139-147.

[164] Schenone, M., Dančík, V., Wagner, B. K., & Clemons, P. A. (2013). Target identification and mechanism of action in chemical biology and drug discovery. Nature chemical biology, 9(4), 232-240.

[165] Friedrich, C. M., & Paterson, T. S. (2004). *In silico* predictions of target clinical efficacy. Drug Discovery Today: Targets, 3(5), 216-222.

[166] Wen, Q. et al. Identification of regulators of polyploidization presents therapeutic targets for treatment of AMKL. Cell 150, 575–589 (2012).

[167] Enna, S. J., & Williams, M. (2009). Challenges in the search for drugs to treat central nervous system disorders. Journal of Pharmacology and Experimental Therapeutics, 329(2), 404-411.

[168] Mei, H., Xia, T., Feng, G., Zhu, J., Lin, S., & Qiu, Y. (2012). Opportunities in systems biology to discover mechanisms and repurpose drugs for CNS diseases. Drug discovery today.

[169] Whittaker P.A. (2004), "The role of bioinformatics in target validation", Drug Discovery Today: Technologies, 1 (2), pp. 125-133.

[170] Jensen, L. J., Saric, J., & Bork, P. (2006). Literature mining for the biologist: from information retrieval to biological discovery. Nature reviews genetics, 7(2), 119-129.

[171] Yang, Y., Adelstein, S. J., & Kassis, A. I. (2009). Target discovery from data mining approaches. Drug discovery today, 14(3), 147-154.

[172] Fernández JM, Hoffmann R, Valencia A. iHOP web services. Nucleic Acids Res. 2007;35:W21-W26.

[173] Plake C, Schiemann T, Pankalla M, Hakenberg J, Leser U. AliBaba: PubMed as a graph. Bioinformatics 2006;22:2444-2445.

[174] Rebholz-Schuhmann D, Kirsch H, Arregui M, Gaudan S, Riethoven M, Stoehr P. EBIMed–text crunching to gather facts for proteins from Medline. Bioinformatics 2007;23:e237-e244.

[175] McEntyre JR, Ananiadou S, Andrews S, Black WJ, Boulderstone R, Buttery P, Chaplin D, Chevuru S, Cobley N, Coleman L-A, et al. UKPMC: a full text article resource for the life sciences. Nucleic Acids Res. 2011;39:D58-D65.

[176] Rebholz-Schuhmann D, Arregui M, Gaudan S, Kirsch H, Jimeno A. Text processing through Web services: calling Whatizit. Bioinformatics 2008;24:296-298.

[177] Doms A, Schroeder M. GoPubMed: exploring PubMed with the Gene Ontology. Nucleic Acids Res. 2005;33:W783-W786.

[178] Thomas, P., Starlinger, J., Vowinkel, A., Arzt, S., & Leser, U. (2012). GeneView: a comprehensive semantic search engine for PubMed. Nucleic acids research, 40(W1), W585-W591.

[179] Hanisch D, Fundel K, Mevissen HT, Zimmer R, Fluck J: ProMiner: rule based protein and gene entity recognition. BMC Bioinforma 2005, 6(Suppl 1):S14

[180] Kim JJ, Pezik P, and Rebholz-Schuhmann D. MedEvi: retrieving textual evidence of relations between biomedical concepts from Medline. Bioinformatics. 2008;24:1410–1412.

[181] Rebholz-Schuhmann D, Kirsch H, Arregui M, et al. EBIMed–text crunching to gather facts for proteins from Medline. Bioinformatics. 2007;23:e237

[182] CiteXplore. 2006. http://www.ebi.ac.uk/citexplore/

[183] Ohta T, Tsuruoka Y, Takeuchi J, et al. Proceedings of the COLING/ACL on Interactive presentation sessions. Sydney, Australia: Association for Computational Linguistics; 2006. An intelligent search engine and GUI-based efficient MEDLINE search tool based on deep syntactic parsing.

[184] Douglas S, Montelione G, Gerstein M. PubNet: a flexible system for visualizing literature derived networks. Genome Biol. 2008. 2005;9:S1.

[185] Wimalasuriya, D. C., & Dou, D. (2010). Ontology-based information extraction: An introduction and a survey of current approaches. Journal of Information Science, 36(3), 306-323.

[186] Friedrich CM, Dach H, Gattermayer T, Engelbrecht G, Benkner S, Hofmann-Apitius M: @neuLink: a service-oriented application for biomedical knowledge discovery. In Proceedings of HealthGrid 2008; 2–4 June 2008. Edited by Solomonides T, Silverstein JC, Saltz J, Legre Y, Kratz M, Foster I, Breton V, Beck JR. IOS Press, Chicago; 2008::165-172.

[187] Klingström, T., & Plewczynski, D. (2011). Protein–protein interaction and pathway databases, a graphical review. Briefings in bioinformatics, 12(6), 702-713.

[188] Mathivanan, S., Periaswamy, B., Gandhi, T. K. B., Kandasamy, K., Suresh, S., Mohmood, R., ... & Pandey, A. (2006). An evaluation of human protein-protein interaction data in the public domain. BMC bioinformatics, 7(Suppl 5), S19.

[189] Garcia-Garcia, J., Guney, E., Aragues, R., Planas-Iglesias, J., & Oliva, B. (2010). Biana: a software framework for compiling biological interactions and analyzing networks. BMC bioinformatics, 11(1), 56.

[190] Kohl, M., Wiese, S., & Warscheid, B. (2011). Cytoscape: software for visualization and analysis of biological networks. In Data Mining in Proteomics (pp. 291-303). Humana Press.

[191] Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. Brief Bioinformatics 2008; 9:75-90.

[192] Héja G, Surján G, Varga P: Ontological analysis of SNOMED CT. BMC Med. Inform. Decis. 2008;8(Suppl. 1):S8.

[193] Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res. 2003;31:334-341.

[194] Whetzel PL, Parkinson H, Causton HC, Fan L, Foster J, Fragoso G, et al. The MGED ontology: a resource for semantics-based description of microarray experiments. Bioinformatics 2006;22:866-873.

[195] Spasic I, Ananiadou S, McNaught J, Kumar A.Text mining and ontologies in biomedicine: making sense of raw text. Brief Bioinformatics 2005;6:239-251.

[196] Doms A, Schroeder M. GoPubMed: exploring PubMed with the gene ontology; Nucleic Acids Res. 2005. p. 33. W783–W786.

[197] Broekstra J, et al. Proceedings of the WWW'04 workshop on Application Design, Development and Implementation Issues in the Semantic Web. New York: 2004. The drug ontology project for Elsevier.

[198] de Bono, B., Grenon, P., & Sammut, S. J. (2012). ApiNATOMY: A novel toolkit for visualizing multiscale anatomy schematics with phenotype‐related information. Human mutation, 33(5), 837-848.

[199] Gómez-Pérez A, et al. Ontological Engineering. Springer, Berlin: 2004.

[200] Gangemi, A., Catenacci, C., Ciaramita, M., & Lehmann, J. (2006). Modelling ontology evaluation and validation. In The Semantic Web: Research and Applications (pp. 140-154). Springer Berlin Heidelberg.

[201] Ivchenko, O., Younesi, E., Shahid, M., Wolf, A., Müller, B., & Hofmann-Apitius, M. (2011). PLIO: an ontology for formal description of protein–ligand interactions. Bioinformatics, 27(12), 1684-1690.

[202] Côté RG, Jones P, Apweiler R, Hermjakob H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics. 2006؛ 7:97.

[203] Grenon P, et al. Biodynamic ontology: applying BFO in the biomedical domain. In: Pisanelli DM, editor. Ontologies in Medicine. Amsterdam: IOS Press; 2004. pp. 20–38.

[204] Younesi, E., Toldo, L., Müller, B., Friedrich, C. M., Novac, N., Scheer, A., ... & Fluck, J. (2012). Mining biomarker information in biomedical literature. BMC medical informatics and decision making, 12(1), 148.

[205] Bug, W. J., Ascoli, G. A., Grethe, J. S., Gupta, A., Fennema-Notestine, C., Laird, A. R., ... & Martone, M. E. (2008). The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. Neuroinformatics, 6(3), 175-194.

[206] Bowden, D. M., Song, E., Kosheleva, J., & Dubach, M. F. (2012). NeuroNames: an ontology for the BrainInfo portal to neuroscience on the web. Neuroinformatics, 10(1), 97-114.

[207] Streit, W.J., Microglia and Alzheimer's disease pathogenesis. J Neurosci Res, 2004. 77(1): p. 1-8.

[208] Gündel, M., Younesi, E., Malhotra, A., Wang, J., Li, H., Zhang, B., ... & Hofmann-Apitius, M. (2013). HuPSON: the

human physiology simulation ontology. Journal of biomedical semantics, 4(1), 35.

[209] Kohl, P., & Noble, D. (2009). Systems biology and the virtual physiological human. Molecular Systems Biology, 5(1).

[210] Lloyd, C. M., Lawson, J. R., Hunter, P. J., & Nielsen, P. F. (2008). The CellML model repository. Bioinformatics, 24(18), 2122-2123.

[211] Malhotra, A., Younesi, E., Gündel, M., Müller, B., Heneka, M. T., & Hofmann-Apitius, M. (2013). ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease. Alzheimer's & Dementia. In Press.

[212] Yu, T. (2010). An exploratory data analysis method to reveal modular latent structures in high-throughput data. BMC bioinformatics, 11(1), 440.

[213] Kelder, T., Conklin, B. R., Evelo, C. T., & Pico, A. R. (2010). Finding the right questions: exploratory pathway analysis to enhance biological discovery in large datasets. PLoS biology, 8(8), e1000472.

[214] Taylor, J. P., Hardy, J., & Fischbeck, K. H. (2002). Toxic proteins in neurodegenerative disease. Science, 296(5575), 1991-1995.

[215] DeCaprio A.P. (Ed.) 2006. Toxicologic biomarkers. Informa Healthcare.

[216] Gashaw, I., Ellinghaus, P., Sommer, A., & Asadullah, K. (2012). What makes a good drug target?. Drug discovery today, 17, S24-S30.

[217] Guan, Y., Gorenshteyn, D., Burmeister, M., Wong, A. K., Schimenti, J. C., Handel, M. A., ... & Troyanskaya, O. G. (2012). Tissue-specific functional networks for prioritizing phenotype and disease genes. PLoS computational biology, 8(9), e1002694.

[218] Bossi A, Lehner B (2009) Tissue specificity and the human protein interaction network. Mol Sys Biol 5: 260.

[219] Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., & Tanzi, R. E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nature genetics, 39(1), 17-23.

[220] Dennis Jr, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: database for annotation, visualization, and integrated discovery. Genome Biol, 4(5), P3.

[221] Almeida OP, NT Lautenschlager (2005) Dementia associated with infectious diseases. International Psychogeriatrics, 17:S65-S77.

[222] Menninger KA. (1926) An analysis of post-influenzal dementia precox as of 1918 and five years later, Am J Psychiatry 82:469-529.

[223] Mattson MP. (2003) Infectious agents and age-related neurodegenerative disorders. Aging Research Reviews, 3(2): 249.

[224] Uldall KK, Marra C. HIV-Associated Dementia, in The HIV Manual, ed. D Spach and T Hooton, Oxford University Press, New York, 1996, pp. 365-368.

[225] Bottiggi K, Chang J, Schmitt F, Avison M, Mootoor Y, Nath A, Berger J (2007) The HIV Dementia Scale: Predictive power in mild dementia and HAART. J Neurol Sci, 260 (1-2): 11-15.

[226] Price RW (1996) Neurological complications of HIV infection. Lancet 348:445-452.

[227] Jang H et al., (2009) Highly pathogenic H5N1 influenza virus can enter the central nervous system and induce neuro inflammation and neurodegeneration. PNAS 33: 14063-14068.

[228] Speth C, Dierich MP, and Sopper S (2005) HIV-infection of the central nervous system: the tightrope walk of innate immunity. Mol Immunol 42: 213-228.

[229] Blasko I, Stampfer-Kountchev M, Robatscher P, Veerhuis R, Eikelenboom P, Grubeck-Loebenstein B (2004) How

chronic inflammation can affect the brain and support the development of Alzheimer's disease in old age: the role of microglia and astrocytes. Aging Cell 3:169-176.

[230] Gao HM et al. (2002) Microglial activation-mediated delayed and progressive degeneration of rat nigral dopaminergic neurons: relevance to Parkinson's disease. J. Neurochem. 81:1285-1297.

[231] Ling et al. (2006) Progressive dopamine neuron loss following supra-nigral lipopolysaccharide (LPS) infusion into rats exposed to LPS prenatally. Exp. Neurol. 199:499-512.

[232] McGeer PL, Schwab C, Parent A, Doudet D (2003) Presence of reactive microglia in monkey substantia nigra years after 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine administration, Ann. Neurol. 54: 599-604.

[233] Wyss-Coray T, Mucke L. 2002. Inflammation in neurodegenerative disease - a double-edged sword. Neuron 35:419-432.

[234] Luca SM, NJ Rothwell, RM Gibson (2006) The role of inflammation in CNS injury and disease. British Journal of Pharmacology 147:S232-S240.

[235] Ghafouri M, S Amini, K Khalili, BE Sawaya. (2006) HIV-1 associated dementia: symptoms and causes. Retrovirology 3:28.

[236] Editorial. (2010) Learning from failure. *Nature Reviews* Drug Discovery 9:499.

[237] Castellani R et al. (2009) Reexamining alzheimer's disease: evidence for a protective role for amylod-beta protein precursor and amyloid-beta. J Alzheimers Dis. 18(2):447-452.

[238] Chatraryamontri A, Ceol A, Peluso D, Nardozza A, Panni S, Sacco F, Tinti M, Smolyar A, Castagnoli L, Vidal M, Cusick ME, Cesareni G. (2009) VirusMINT: a viral protein interaction database. Nucleic Acids Res. 37(Database issue):D669-73.

[239] Fury W, F Batliwalla, PK Gregersen, W Li. (2006) Overlapping probabilities of top ranking gene lists,

hypergeometric distribution, and stringency of gene selection criterion. Conf Proc IEEE Eng Med Biol Soc. 1:5531-5534.

[240] Blalock, E. M., Geddes, J. W., Chen, K. C., Porter, N. M., Markesbery, W. R. & Landfield, P. W. (2004) Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. PNAS 2004 101:2173-2178.

[241] Ortutay, C., & Vihinen, M. (2006). Immunome: a reference set of genes and proteins for systems biology of the human immune system. Cellular immunology, 244(2), 87-89.

[242] Online Mendelian Inheritance in Man, OMIM (TM). McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), {date of download: 03.11.2009}. World Wide Web URL: www.ncbi.nlm.nih.gov/omim/.

[243] Sun, H.; Li, D.; Wang, J.; Liu, Z.; Zhu, Y.; Xie, H.; He, F. (2009) PNmerger: a Cytoscape plugin to merge biological pathways and protein interaction networks. Progress in Biochemistry and Biophysics 36(12): 1613-1616.

[244] Itzhaki RF, Wozniak MA, Appelt DM, Balin BJ (2004) Infiltration of the brain by pathogens causes Alzheimer's disease. Neurobiology of Aging 25:619-627.

[245] Loeb MB et al. (2004) A randomized controlled trial of Doxycycline and Rifampin for patients with Alzheimer's disease. Journal of the American Geriatrics Society, 52: 381–387.

[246] Kim YS, Joh TH (2006) Microglia, major player in the brain inflammation: their role in the pathogenesis of Parkinson's disease. Exp mol Med. 38(4): 333-47.

[247] Di Leonardo A, Linke SP, Clarkin K, Wahl GM (1996) DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. Genes Dev. 10:934-947.

[248] Guillemin GJ, Cullen KM, Lim CK, Smythe GA, Garner B, Kapoor V, et al. (2007) Characterization of the kynurenine pathway in human neurons. J Neurosci. 27:12884–92.

[249] Hartai Z et al. (2007) Decreased serum and red blood cell kynurenic acid levels in Alzheimer's disease. Neurochem Int. 50(2):308-13.

[250] Heyes MP, Brew BJ, Saito K, Quearry BJ, Price RW, Lee K, Bhalla RB, Der M, Markey SP (1992) Inter-relationships between quinolinic acid, neuroactive kynurenines, neopterin and beta 2-microglobulin in cerebrospinal fluid and serum of HIV-1-infected patients. J Neuroimmunol. 40(1):71-80.

[251] Huh, Y. et al. (2003) Microglial activation and tyrosine hydroxylase immunoreactivity in the substantia nigral region following transient focal ischemia in rats. Neurosci. Lett. 349, 63–67.

[252] McGeer, P. L., Schwab, C., Parent, A. & Doudet, D. (2003) Presence of reactive microglia in monkey substantia nigra years after 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine administration. Ann. Neurol. 54,599–604.

[253] Gao, H. M., Liu, B., Zhang, W. & Hong, J. S. (2003) Critical role of microglial NADPH oxidase-derived free radicals in the in vitro MPTP model of Parkinson's disease. FASEB J. 17, 1954–1956.

[254] Rezaei P, Lantos PL (2001) Microglia and the pathogenesis of spongiform encephalopathies. Brain Res Rev 35: 55-72.

[255] Younesi, E., & Hofmann-Apitius, M. (2013). A network model of genomic hormone interactions underlying dementia and its translational validation through serendipitous off-target effect. Journal of translational medicine, 11(1), 177.

[256] D. J. Brooks, Imaging amyloid in Parkinson's disease dementia and dementia with Lewy bodies with positron emission tomography. Mov. Disord. 24(Suppl 2): S742-7 (2009).

[257] E. Vanmechelen, H. Vanderstichele, F. Hulstaert, N. Andreasen, L. Minthon, B. Winbald, P. Davidsson, K.

Blennow, Cerebrospinal fluid τ and β-amyloid(1-42) in dementia disorders. Mech. Ageing Dev. 122: 2005-2011 (2001).

[258] C. A. Ross, M. A. Poirier, Protein aggregation and neurodegenerative disease. Nat. Med. 10: S10-S17 (2004).

[259] S. K. Maji, M. H. Perrin, M. R. Sawaya, S. Jessberger, K. Vadodaria, R. A. Rissman, P. S. Singru, K. P. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, P. Sawchenko, W. Vale, R. Riek, Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. Science 325: 328–332 (2009).

[260] R. B. Gibbs, Estrogen therapy and cognition: a review of the cholinergic hypothesis. Endocr. Rev. 31: 224-253 (2010).

[261] S. Artero, M. L. Ancelin, F. Portet, A. Dupuy, C. Berr, J. F. Dartigues, C. Tzourio, O. Rouaud, M. Poncet, F. Pasquier, S. Auriacombe, J. Touchon, K. Ritchie, Risk profiles for mild cognitive impairment and progression to dementia are gender specific. J. Neurol. Neurosurg. Psychiatry 79: 979-984 (2008).

[262] G. Ravaglia, P. Forti, F. Maioli, L. Bastagli, F. Montesi, N. Pisacane, M. Chiappelli, F. Licastro, C. Patterson, Endogenous sex hormones as risk factors for dementia in elderly men and women. J. Gerontol. A Biol. Sci. Med. Sci. 62: 1035-1041 (2007).

[263] A. F. Jorm, D. Jolley, The incidence of dementia, a meta analysis. Neurology 51: 728-733 (1998).

[264] A. Paganini-Hill, V. Henderson, Estrogen replacement therapy and risk of Alzheimer disease. Arch. Intern. Med. 156: 2213–2217 (1996).

[265] M. X. Tang, D. Jacobs, Y. Stern, K. Marder, P. Schofield, B. Gurland, H. Andrews, R. Mayeux, Effect of oestrogen during menopause on risk and age at onset of Alzheimer's disease. Lancet 348: 429–432 (1996).

[266] P Zandi, M. Carlson, B. Plassman, K. Welsh-Bohmer, L. Mayer, D. Steffens, J. Breitner, Hormone replacement therapy

and incidence of Alzheimer disease in older women. JAMA-J. Am. Med. Assoc. 288: 2123–2129 (2002).

[267] M. Espeland, S. Rapp, S. Shumaker, R. Brunner, J. Manson, B. Sherwin, J Hsia, K. Margolis, P. Hogan, R. Wallace, M. Dailey, R. Freeman, J. Hays, Conjugated equine estrogens and global cognitive function in postmenopausal women. JAMA-J. Am. Med. Assoc. 291: 2959–2968 (2004).

[268] S. A. Shumaker, C. Legault, S. R. Rapp, L. Thal, R. B. Wallace, J. K. Ockene, S. L. Hendrix, B. N. Jones, 3rd., A. R. Assaf, R. D. Jackson, J. M. Kotchen, S. Wassertheil-Smoller, J. Wactawski-Wende, Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women. The women's health initiatitive memory study: a randomized controlled trial. JAMA-J. Am. Med. Assoc. 289: 2651–2662 (2003).

[269] J. Ryan, I. Carriere, J. Scali, J. F. Dartigues, C. Tzourio, M. Poncet, K. Ritchie, M. L. Ancelin, Characteristics of hormone therapy, cognitive function, and dementia. Neurology 73: 1729-1737 (2009).

[270] W. Lieb, A. S. Beiser, R. S. Vasan, Z. S. Tan, R. Au, T. B. Harris, R. Roubenoff, S. Auerbach, C. DeCarli, P. A. Wolf, S. Seshadri, Association of plasma leptin levels with incident Alzheimer disease and MRI measures of brain aging. JAMA-J. Am. Med. Assoc. 302:2565-2572 (2009).

[271] N. Tezapsidis, J.M. Johnston, M.A. Smith, J.W. Ashford, G. Casadesus, N.K. Robakis, B. Wolozin, G. Perry, X. Zhou, S.J. Greco, S. Sarkar, Leptin: a novel therapeutic strategy for alzheimer's disease. J. Alzheimers Dis. 16: 731-740 (2009).

[272] C. P. Maurizi, Loss of intraventricular fluid melatonin can explain the neuropathology of Alzheimer's disease. Med. Hypotheses 49: 153-158 (1997).

[273] R. Y. Liu, J. N. Zhou, J. van Heerikhuize, M. A. Hofman, D. F. Swaab, Decreased melatonin levels in post-mortem cerebrospinal fluid in relation to aging, Alzheimer's disease,

and apolipoprotein E-epsilon4/4 genotype. J. Clin. Endocrinol. Metab. 84: 323-327 (1999).

[274] K. Mishima, T. Tozawa, K. Satoh, Y. Matsumoto, Y. Hishikawa, M. Okawa, Melatonin secretion rhythm disorders in patients with senile dementia of Alzheimer's type with disturbed sleep-waking. Biol. Psychiatry 45: 417-421 (1999).

[275] E. Atsubara, T. Bryant-Thomas, J. Pacheco Quinto, T. L. Henry, B. Poeggeler, D. Herbert, F. Cruz-Sanchez, Y. J. Chyan, M. A. Smith, G. Perry, M. Shoji, K. Abe, A. Leone, I. Grundke-Ikbal, G. L. Wilson, J. Ghiso, C. Williams, L. M. Refolo, M. A. Pappolla, D. G. Chain, E. Neria, Melatonin increases survival and inhibits oxidative and amyloid pathology in a transgenic model of Alzheimer's disease. J. Neurochem. 85: 1101-1108 (2003).

[276] M. M. Breteler, C. M. van Duijn, V. Chandra, L. Fratiglioni, A. B. Graves, A. Heyman, A. F. Jorm, E. Kokmen, K. Kondo, J. A. Mortimer, W. A. Rocca, S. L. Shalat, H. Soininen, A. Hofman, Medical history and the risk of Alzheimer's disease: a collaborative re-analysis of case-control studies. Int. J. Epidemiol. 20(Suppl 2): S36-S42 (1991).

[277] M. Ganguli, L. A. Burmeister, E. C. Seaberg, S. Belle, S. T. DeKosky, Association between dementia and elevated TSH: a community-based study. Biol. Psychiatry 40: 714-725 (1996).

[278] A. L. Fu, C. Y. Zhou, X. Chen, Thyroid hormone prevents cognitive deficit in a mouse model of Alzheimer's disease. Neuropharmacology 58: 722-729 (2010).

[279] C. May, S. I. Rapoport, T. P. Tomai, G. P. Chrousos, P. W. Gold, Cerebral spinal fluid concentrations of corticotropin-releasing hormone (CRH) and corticotropin (ACTH) are reduced in patients with Alzheimer's disease. Neurology 37: 535-538 (1987).

[280] N. Pomara, R. R. Singh, D. Deptula, P. A. LeWitt, G. Bissette, M. Stanley, C. B. Nemeroff, CSF corticotropin-releasing factor (CRH) in Alzheimer's disease. Its

relationship to severity of dementia and monoamine metabolites. Biol. Psychiatry 26: 500-504 (1989).

[281] A Aleman, H. J. Verhaar, E. H. F. de Haan, W. R. De Vries, M. M. Samson, M. L. Drent, E. A. Van der Veen, H.P. Koppeschaar, Insulin-like growth factor-I and cognitive function in healthy older men. J. Clin. Endocrinol. Metab. 84: 471–5 (1999).

[282] A. Rollero, G. Murialdo, S. Fonzi, S. Garrone, M. V. Gianelli, E. Gazzerro, A. Barreca, A. Polleri, Relationship between cognitive function, growth hormone and insulin-like growth factor I plasma levels in aged subjects. Neuropsychobiology 23: 45–55 (1998).

[283] B. R. Lackey, S. L. Gray, D. M. Henricks, Actions and interactions of the IGF system in Alzheimer's Disease: review and hypotheses. Growth Horm. IGF Res. 10: 1–13 (2000).

[284] M. V. Vitiello, K. E. Moe, G. R. Merriam, G. Mazzoni, D. H. Buchner, R. S. Schwartz, Growth hormone releasing hormone improves the cognition of healthy older adults. Neurobiol. Aging 27: 318-323 (2006).

[285] D. Y. Chen, S. A. Stern, A. Garcia-Osta, B. Saunier-Rebori, G. Pollonini, D. Bambah-Mukku, R. D. Blitzer, C. M. Alberini, A critical role for IGF-II in memory consolidation and enhancement. Nature 469: 491–497 (2011).

[286] S. Craft S, E. Peskind, M. W. Schwartz, G. D. Schellenberg, M. Raskind, D. Porte, Jr., Cerebrospinal fluid and plasma insulin levels in Alzheimer's disease: relationship to severity of dementia and apolipoprotein E genotype. Neurology 50: 164-168 (1998).

[287] S. Craft, S. Asthana, D. G. Cook, L. D. Baker, M. Cherrier, K. Purganan, C. Wait, A. Petrova, S. Latendresse, G. S. Watson, J. W. Newcomer, G. D. Schellenberg, A. J. Krohn, Insulin dose-response effects on memory and plasma amyloid precursor protein in Alzheimer's disease: interactions with apolipoprotein E genotype. Psychoneuroendocrino. 28: 809-822 (2003).

[288] A. M. Moloney, R. J. Griffin, S. Timmons, R. O'Connor, R. Ravid, C. O'Neill, Defects in IGF-1 receptor, insulin receptor and IRS-1/2 in Alzheimer's disease indicate possible resistance to IGF-1 and insulin signaling. Neurobiol. Aging 31: 224-243 (2010).

[289] L. D. Baker, D. J. Cross, S. Minoshima, D. Belongia, G. Stennis Watson, S. Craft, Insulin Resistance and Alzheimer-like Reductions in Regional Cerebral Glucose Metabolism for Cognitively Normal Adults With Prediabetes or Early Type 2 Diabetes. Arch. Neurol. 68: 51-57 (2011).

[290] C. Sims-Robinson, B. Kim, A. Rosko, E. L. Feldman, How does diabetes accelerate Alzheimer disease pathology? Nat. Rev. Neurol. 6: 551-559 (2010).

[291] A. Ott, R. P. Stolk, F. van Harskamp, D. E. Grobbee, M. M. Breteler, Diabetes mellitus and the risk of dementia: The Rotterdam Study. Neurology 53: 1937-1942 (1999).

[292] S. M. de la Monte, J. R. Wands, Alzheimer's disease is type 3 diabetes-evidence reviewed. J. Diabetes Sci. Technol. 2: 1101–1113 (2008).

[293] K. Akter, E. A. Lanza, S. A. Martin, N. Myronyuk, M. Rua, R. B. Raffa, Diabetes mellitus and Alzheimer's disease: shared pathology and treatment? Brit. J. Clin. Pharmaco. 71: 365–376 (2011).

[294] N. Saleh, S. Moutereau, A. Durr, P. Krystkowiak, J. P. Azulay, C. Tranchant, E. Broussolle, F. Morin, A. C. Bachoud-Lévi, P. Maison, Neuroendocrine disturbances in huntington's disease. PLoS One 4(3): e4962 (2009).

[295] A. Califano, A. J. Butte, S. Friend, T. Ideker, E. Schadt, Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nat. Genet. 44: 841-847 (2012).

[296] A. Potapov, I. Liebich, J. Dönitz, K. Schwarzer, N. Sasse, T. Schoeps, T. Crass, E. Wingender, EndoNet: an information resource about endocrine networks. Nucleic Acids Res. 34: D540-D545 (2006).

[297] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks. (International AAAI Conference on Weblogs and Social Media, San Jose, California, 2009).

[298] G. Su, A. Kuchinsky, J. H. Morris, D. J. States, F. Meng, GLay: community structure analysis of biological networks. Bioinformatics 26: 3135-3137 (2010).

[299] D. Nishimura, BioCarta. Biotech. Softw. Internet Rep. 2: 117-120 (2011)

[300] G. Frühbeck, Intracellulat signalling pathways activated by leptin. Biochem. J. 393: 7-20 (2006).

[301] L. C. Moeller, X. Cao, A. M. Dumitrescu, H. Seo, S. Refetoff, Thyroid hormone mediated changes in gene expression can be initiated by cytosolic action of the thyroid hormone receptor β through the phosphatidylinositol 3-kinase pathway. Nucl. Recept. Signal. 4:e020 (2006).

[302] https://www.qiagen.com/geneglobe/pathwayview.aspx?pathwayID=124

[303] M. L. Dubocovich, M. A. Rivera-Bermudez, M. J. Gerdin, M. I. Masana, Molecular pharmacology, regulation and function of mammalian melatonin receptors. Front. Biosci. 8: d1093-108 (2003).

[304] S. E. Arnold, B. T. Hyman, J. Flory, The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in cerebral cortex of patients with Alzheimer's disease. Cereb. Cortex 1: 103–116 (1991).

[305] Davis JD, Podolanczuk A, Donahue JE, Stopa E, Hennessey JV, Luo LG, Lim YP, Stern RA: Thyroid hormone levels in the prefrontal cortex of post-mortem brains of Alzheimer's disease patients.

[306] Dai J, Buijs R, Swaab D: Glucocorticoidx hormone (cortisol) affects axonal transport in human cortex neurons but shows resistance in Alzheimer's disease. Br J Pharmacol 2004, 143:606-610.

[307] Souza EB, Whitehouse PJ, Price DL, Vale WW: Abnormalities in Corticotropin-releasing Hormone (CRH) in Alzheimer's disease and other human disorders. Ann N Y Acad Sci 1987, 512:237-247. Curr Aging Sci 2008, 1:175.

[308] Ishunina TA, Kamphorst W, Swaab DF: Metabolic alterations in the hypothalamus and basal forebrain in vascular dementia. J Neuropathol Exp Neurol 2004, 63:1243-1254.

[309] Murialdo G, Zerbi F, Filippi U, Tosca P, Fonzi S, Di Paolo E, Costelli P, Porro S, Polleri A, Savoldi F: Cholinergic modulation of growth hormone-releasing hormone effects on growth hormone secretion in dementia. Neuropsychobiology 2008, 24:129-134.

[310] Keenan PA, Ezzat WH, Ginsburg K, Moore GJ: Prefrontal cortex as the site of estrogen's effect on cognition. Psychoneuroendocrino 2001, 26:577-590.

[311] Watson GS, Bernhardt T, Reger MA, Cholerton BA, Baker D, Peskin ER, Asthana S, Plymate SR, Frölich L, Craft S: Insulin effects on CSF norepinephrine and cognition in Alzheimer's disease. Neurobiol Aging 2006, 27:38-41.

[312] Arwert LI, Veltman DJ, Deijen JB, van Dam PS, Drent ML: Effects of growth hormone substitution therapy on cognitive functioning in growth hormone deficient patients: a functional MRI study. Neuroendocrinology 2006, 83:12-19.

[313] Zhu DF, Wang ZX, Zhang DR, Pan ZL, He S, Hu XP, Chen XC, Zhou JN: fMRI revealed neural substrate for reversible working memory dysfunction in subclinical hypothyroidism. Brain 2006, 129:2923-2930.

[314] Liang WS, Dunckley T, Beach TG, Grover A, Mastroeni D, Ramsey K, Caselli RJ, Kukull WA, McKeel D, Morris JC, Hulette CM, Schmechel D, Reiman EM, Rogers J, Stephan DA: Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. Physiol Genomics 2008, 33:240-256.

[315] Liang D, Han G, Feng X, Sun J, Duan Y, Lei H: Concerted perturbation observed in a hub network in Alzheimer's disease. PLoS One 2012, 7:e40498.

[316] MacDonald ML, Lamerdin J, Owens S, Keon BH, Bilter GK, Shang Z, Huang Z, Yu H, Dias J, Minami T, Michnick SW, Westwick JK: Identifying off-target effects and hidden phenotypes of drugs in human cells. Nat Chem Biol 2006, 2:329-337.

[317] Singh M, Sétáló G, Guan X, Warren M, Toran-Allerand CD: Estrogen-induced activation of mitogen-activated protein kinase in cerebral cortical explants: convergence of estrogen and neurotrophin signaling pathways. J Neurosci 1999, 19:1179-1188.

[318] Craft S, Asthana S, Schellenberg G, Cherrier M, Baker LD, Newcomer J, Plymate S, Latendresse S, Petrova A, Raskind M, Peskind E, Lofgreen C, Grimwood K: Insulin metabolism in Alzheimer's disease differs according to apolipoprotein E genotype and gender. Neuroendocrino 1999, 70:146-152.

[319] Cardona-Gomez P, Perez M, Avila J, Garcia-Segura LM, Wandosell F: Estradiol inhibits GSK3 and regulates interaction of estrogen receptors, GSK3, and β-catenin in the hippocampus. Mol Cell Neurosci 2004, 25:363-373.

[320] Cuadrado A, Garcia-Fernandez LF, Imai T, Okano H, Munoz A: Regulation of tau RNA maturation by thyroid hormone is mediated by the neural RNA-binding protein Musashi-1. Mol Cell Neurosci 2002, 20:198-210.

[321] Wang JZ, Wang ZF: Role of melatonin in Alzheimer-like neurodegeneration. Acta Pharmacol Sin 2006, 27:41-49.

[322] Rissman RA, Lee KF, Vale W, Sawchenko PE: Corticotropin-releasing factor receptors differentially regulate stress-induced τ phosphorylation. J Neurosci 2007, 27:6552-6562.

[323] Ohara T, Doi Y, Ninomiya T, Hirakawa Y, Hata J, Iwaki T, Kanba S, Kiyohara Y: Glucose tolerance status and risk of

dementia in the community: the Hisayama study. Neurology 2011, 77:1126-1134.

[324] Hsu CC, Wahlqvist ML, Lee MS, Tsai HN: Incidence of Dementia is Increased in Type 2 Diabetes and Reduced by the Use of Sulfonylureas and Metformin. J Alzheimers Dis 2011, 24:485-493.

[325] Xu W, Qiu C, Gatz M, Pedersen NL, Johansson B, Fratiglioni L: Mid- and Late-life diabetes in relation to the risk of dementia: a population-based twin study. Diabetes 2009, 58:71-77.

[326] Yaffe K, Falvey C, Hamilton N, Schwartz AV, Simonsick EM, Satterfield S, Cauley JA, Rosano C, Launer LJ, Strotmeyer ES, Harris TB: Diabetes, glucose control, and 9-year cognitive decline among older adults without dementia. Arch Neurol 2012, 9:1170-1175.

[327] Cammarota M, Bevilaqua LRM, Ardenghi P, Paratcha G, Levi de Stein M, Izquierdo I, Medina JH: Learning-associated activation of nuclear MAPK, CREB and Elk-1, along with Fos production, in the rat hippocampus after a one-trial avoidance learning: abolition by NMDA receptor blockade. Mol Brain Res 2000, 76:36-46.

[328] Zhao WQ, Alkon DL: Role of insulin and insulin receptor in learning and memory. Mol Cell Endocrinol 2001, 177:125-134.

[329] Zhao WQ, Chen H, Quon MJ, Alkon DL: Insulin and the insulin receptor in experimental models of learning and memory. Eur J Pharmacol 2004, 490:71-81.

[330] Jucker M: The benefits and limitations of animal models for translational research in neurodegenerative diseases. Nat Med 2010, 16:1210-1214.

[331] Hölscher C: Neuronal mechanisms of memory formation: concepts of long term potentiation and beyond. Cambridge: Cambridge University Press; 2001.

[332] Lynch MA: Long-term potentiation and memory. Physiol Rev 2004, 84:87-136.

[333] Buxbaum JD, Cullen EI, Friedhoff LT: Pharmacological concentrations of the HMG-CoA reductase inhibitor lovastatin decrease the formation of the Alzheimer beta-amyloid peptide in vitro and in patients. Front Biosci 2002, 7:a50-a59.

[334] Banks WA: Brain meets body: the blood–brain barrier as an endocrine interface. Endocrinology 2012, 9:4111-411.

[335] Younesi, E., & Hofmann-Apitius, M. (2013). Biomarker-guided translation of brain imaging into disease pathway models. Scientific reports, 3. [http://www.nature.com/srep/2013/131129/srep03375/full/srep03375.html]

[336] Kherlopian, A. R. *et al*. A review of imaging techniques for systems biology. *BMC Syst. Biol.* **2**, 74 (2008).

[337] Monchi, O., Benali, H., Doyon, J. & Strafella, A. P. Recent advances in neuroimaging methods. *Int. J. Biomed. Imaging* 2008, 1-2 (2008).

[338] Johnson, K. A., Fox, N. C., Sperling, R. A. & Klunk, W. E. Brain imaging in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2**, a006213 (2012).

[339] Mattson, M. P. Pathways towards and away from Alzheimer's disease. *Nature* 430, 631-639 (2004).

[340] Ashburner, J. *et al*. Computer-assisted imaging to assess brain structure in healthy and diseased brains· *Lancet Neurol.* 2, 79–88 (2003).

[341] Coleman, M. R. *et al*. Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain* 132, 2541-2552 (2009).

[342] Bansal, R. *et al*. Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. *PLoS ONE* 7, e50698 (2012).

[343] de Backer, M. E. *et al*. MR-based molecular imaging of the brain: the next frontier. *Am. J. Neuroradiol.* 31, 1577-1583 (2010).

[344] Baskys, A. & Blaabjerg, M. Understanding regulation of nerve cell death by mGluRs as a method for development of successful neuroprotective strategies. J. Neurol. Sci. 229, 201–209 (2005).

[345] Strasser, A., Jost, P. J. & Nagata, S. The many roles of FAS receptor signaling in the immune system. Immunity 30, 180–192 (2009).

[346] Kim, E. K. & Choi, E. J. Pathological roles of MAPK signaling pathways in human diseases. Bioch. Biophys. Acta 1802, 396–405 (2010).

[347] León, D., & Markel, S. (Eds.). (2010). In Silico Technologies in Drug Target Identification and Validation. CRC Press.

[348] Drews J. Drug discovery: a historical perspective. Science 2000, 287:1960-1964.

[349] Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? Nat Rev Drug Discov. 2006, 5:993-996.

[350] Hopkins, A. L., & Groom, C. R. (2003). Target analysis: a priori assessment of druggability. In Small Molecule—Protein Interactions (pp. 11-17). Springer Berlin Heidelberg.

[351] Cheng, A. C., Coleman, R. G., Smyth, K. T., Cao, Q., Soulard, P., Caffrey, D. R., ... & Huang, E. S. (2007). Structure-based maximal affinity model predicts small-molecule druggability. Nature biotechnology, 25(1), 71-75.

[352] Owens, J. (2007). Determining druggability. Nature Reviews Drug Discovery, 6(3), 187-187.

[353] Lindsay, M. A. (2005). Finding new drug targets in the 21st century. Drug discovery today, 10(23), 1683-1687.

[354] Babic, I., & Mischel, P. S. (2012). Discovery in Context: Leveraging Multidimensional Glioblastoma Datasets to Identify Targetable Regulatory Networks. Cancer Discovery, 2(8), 676-678.

[355] Balbin, O. A., Prensner, J. R., Sahu, A., Yocum, A., Shankar, S., Malik, R., ... & Chinnaiyan, A. M. (2013).

Reconstructing targetable pathways in lung cancer by integrating diverse omics data. Nature communications, 4.

[356] Mahajan, R., & Gupta, K. (2010). Food and drug administration's critical path initiative and innovations in drug development paradigm: Challenges, progress, and controversies. Journal of Pharmacy And Bioallied Sciences, 2(4), 307.

[357] Hellerstein, M. K. (2008). A critique of the molecular target-based drug discovery paradigm based on principles of metabolic control: Advantages of pathway-based discovery. Metabolic engineering, 10(1), 1-9.

[358] Shankaran, M., King, C., Lee, J., Busch, R., Wolff, M., Hellerstein, M.K., 2006. Discovery of novel hippocampal neurogenic agents by using an in vivo stable isotope labeling technique. J. Pharmacol. Exp. Ther. 319 (3), 1172–1181.

[359] Shankaran, M., Marino, M., Busch, R., Keim, C., King, C., Lee, J., Killion, S., Awada, M., Hellerstein, M., 2007. Measurement of brain microglial proliferation rates in vivo in response to neuroinflammatory stimuli: application to drug discovery. J. Neurosci. Res. 85 (11), 2374–2384.

[360] Cramer, P. E., Cirrito, J. R., Wesson, D. W., Lee, C. D., Karlo, J. C., Zinn, A. E., ... & Landreth, G. E. (2012). ApoE-directed therapeutics rapidly clear β-amyloid and reverse deficits in AD mouse models. Science, 335(6075), 1503-1506.

[361] Li, J., & Lu, Z. (2013). Pathway-based drug repositioning using causal inference. BMC Bioinformatics, 14(Suppl 16), S3.

[362] Haddad, J. J., Saadé, N. E., & Safieh-Garabedian, B. (2002). Cytokines and neuro–immune–endocrine interactions: a role for the hypothalamic–pituitary–adrenal revolving axis. Journal of neuroimmunology, 133(1), 1-19.

[363] Kim, E. K., & Choi, E. J. (2010). Pathological roles of MAPK signaling pathways in human diseases. Biochimica et

Biophysica Acta (BBA)-Molecular Basis of Disease, 1802(4), 396-405.

[364] Dineley, K. T., Westerman, M., Bui, D., Bell, K., Ashe, K. H., & Sweatt, J. D. (2001). β-Amyloid activates the mitogen-activated protein kinase cascade via hippocampal α7 nicotinic acetylcholine receptors: in vitro and in vivo mechanisms related to Alzheimer's disease. The Journal of Neuroscience, 21(12), 4125-4133.

[365] Moh, C., Kubiak, J. Z., Bajic, V. P., Zhu, X., Smith, M. A., & Lee, H. G. (2011). Cell cycle deregulation in the neurons of Alzheimer's disease. In Cell Cycle in Development (pp. 565-576). Springer Berlin Heidelberg.

[366] Kaminska, B., Gozdz, A., Zawadzka, M., Ellert-Miklaszewska, A., & Lipko, M. (2009). MAPK signal transduction underlying brain inflammation and gliosis as therapeutic target. The Anatomical Record, 292(12), 1902-1913.

[367] Chico, L. K., Van Eldik, L. J., & Watterson, D. M. (2009). Targeting protein kinases in central nervous system disorders. Nature Reviews Drug Discovery, 8(11), 892-909.

[368] Piau, A., Nourhashémi, F., Hein, C., Caillaud, C., & Vellas, B. (2011). Progress in the development of new drugs in Alzheimer's disease. The journal of nutrition, health & aging, 15(1), 45-57.

[369] Björkqvist, M., Ohlsson, M., Minthon, L., & Hansson, O. (2012). Evaluation of a previously suggested plasma biomarker panel to identify Alzheimer's disease. PloS one, 7(1), e29868.

[370] Catlett, N. L., Bargnesi, A. J., Ungerer, S., Seagaran, T., Ladd, W., Elliston, K. O., & Pratt, D. (2013). Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data. BMC bioinformatics, 14(1), 340.