Building multiscale computable model of Alzheimer's disease and identification of novel mechanisms for new therapeutic interventions

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Alpha Tom Kodamullil

aus

Kerala, Indien

Bonn, 2018

Angefertigt mit Genehmigung

der Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

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

2. Gutachter: **Prof. Dr. Heiko Schoof**

Tag der Promotions: June 5th 2018

Erscheinungsjahr: 2018

Abstract

Despite an avalanche of data in the field of biomedicine, we are obviously not managing well to extract meaningful information from this vast amount of data to better understand complex diseases and their mechanisms. Something must be wrong with the paradigmatic “let us generate new data”, that drives current biomedical research. After all, we still have only a limited number of approved drugs available for many complex diseases; interpreting data and associating them with underlying molecular mechanisms of the disease is still a substantial challenge. Approaches that look into a wider perspective of the whole disease etiology as opposed to investigating on specific perturbed pathways or differentially expressed genes bear the potential to go beyond mere pattern identification. Biological networks help to achieve this goal acting as a platform to integrate heterogeneous data and a priori knowledge that may comprise various causal and correlational relationships among biological entities. These networks will lay the ground for the identification of disease mechanisms.

This thesis presents a new formalism that integrate all combinations of interactions with various types of entities from different sources to understand how a single perturbance between two interactors can totally modify or amplify the changes of the whole system. As a use case, I have built the biggest computable mechanistic model of Alzheimer’s disease (AD) in the course of this work. The first outcome is the identification of an early perturbed mechanism on AD based on interference with the neurotrophin signaling pathway. Secondly, I have linked SNP-associated effects to a larger functional context, which corroborates the comorbid association between AD and type 2 diabetes mellitus. Thirdly, I have systematically linked genetic and epigenetic alterations of DNA to the

aetiology of diseases. Whilst the established computable model is specific to human pathophysiology, I have taken the opportunity of its existence to tackle one of the key questions of translational Alzheimer research, namely the functional equivalence of transgenic mouse models with the human disease pathophysiology. I compared the functional, mechanism inventory of a pre-clinical mouse model with the pathophysiology mechanisms that were described for humans in the area of neuro-inflammation. That analysis was extended towards pharmacology, where I analyzed – on the basis of the putative mechanism of action of a discontinued AD targeted drug; Celecoxib, - the reasons why that drug failed in the late phases of clinical trials. As I could show, the pre-clinical mouse experiments did not reflect the mechanistic context that is active in humans; which explains at mechanism-level the late failure of the drug despite promising results of the pre-clinical studies done with experimental animals. Lastly, I have used a comprehensive inventory of Alzheimer disease mechanisms to trace the investment of the pharmaceutical industry in AD drug development. I could demonstrate, how small the spectrum of candidate pathophysiology mechanisms is that the pharmaceutical industry is working on and I could show, how reluctant big pharma companies are to move from the “established targets” or “well-known pathways” into mechanisms that are novel, “ignored” or at least “not targeted” yet.

Declaration

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

Signature: ……..

Author: Alpha Tom Kodamullil

Acknowledgement

First, I would like to express my sincere gratitude to Prof. Dr. Martin Hofmann-Apitius for giving me the opportunity to do my PhD thesis under his supervision and his valuable support and guidance during my PhD.

I would like to thank Prof. Dr. Heiko Schoof for his valuable review. It is an honor to have you as the second reviewer for my thesis. My sincere gratitude also goes to Prof. Dr. Andreas Weber and Prof. Dr. Diana Imhof for agreeing to be part of PhD committee.

I would like to express my sincere thanks to Dr. Erfan Younesi for his valuable support, patience, motivation and guidance as an internal supervisor. My sincere thanks goes to Dr. Allison Provost and Sarah Mubeen for reviewing the PhD thesis.

I would like to thank my master thesis students Reagon Karki, Mohammad Asif Emon, Afroza Khanam Irin and Daniel Domingo-Fernández for contributing to good scientific discussions. Furthermore, my sincere thanks goes to all my students who worked with me for doing the curation task.

I thank to our colleagues from Fraunhofer SCAI, especially Stephan Springstubbe for creating a wonderful work environment.

Last but not the least, I would like to thank my family and friends for always being supportive and encouraging throughout my studies.

Data does not equal information; information does not equal knowledge; and, most importantly of all, knowledge does not equal wisdom. We have oceans of data, rivers of information, small puddles of knowledge, and the odd drop of wisdom.

- Henry Nix, Keynote address, AURISA, 1990

Publications

- **Alpha Tom Kodamullil**, Firas Zekri, Meemansa Sood, Bastian Hengerer, Luc Canard, Duncan McHale, Martin Hofmann-Apitius. 'Tracing the money for development of AD drugs: mapping investment to opportunities and possible drug target mechanisms'
  Nature Reviews Drug Discovery (2017): 819
- **Alpha Tom Kodamullil**, Anandhi Iyappan, Reagon Karki, Sumit Madan, Mufassra Naz, Erfan Younesi, Martin Hofmann-Apitius. 'Of mice and men: comparative analysis of neuro-inflammatory mechanisms in human and mouse using cause and-effect models'
  Journal of Alzheimer's disease (2017):1045-1055
- **Alpha Tom Kodamullil**, Erfan Younesi, Mufassra Naz, Shweta Bagewadi, Martin Hofmann-Apitius. 'Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis'
  Alzheimer's & Dementia (2015):1329-39
- Afroza Khanam Irin, **Alpha Tom Kodamullil**, Michaela Gündel, Martin Hofmann-Apitius. 'Computational modeling approaches on epigenetic factors in neurodegenerative and autoimmune diseases and their mechanistic analysis'
  Journal of Immunology Research (2015):737168
- Mufassra Naz, **Alpha Tom Kodamullil**, Martin Hofmann-Apitius. 'Reasoning over Genetic Variance Information in Cause-and-Effect Models of Neurodegenerative Diseases'

Briefings in Bioinformatics (2016): 505-516

- Mohammad Asif Emran Khan Emon, **Alpha Tom Kodamullil**, Erfan Younesi, Reagon Karki, Martin Hofmann-Apitius. 'Using Drugs as Molecular Probes: Enabling a Chemical Biology Approach in Molecular Systems Biology of the Alzheimer's Disease'

  Journal of Alzheimer's disease (2017):677-686

- Daniel Domingo-Fernández, **Alpha Tom Kodamullil**, Anandhi Iyappan, Mufassra Naz, Mohammad Asif Emran Khan Emon, Tamara Raschka, Reagon Karki, Stephan Springstubbe, Christian Ebeling, Martin Hofmann-Apitius. 'Multimodal Mechanistic Signatures for Neurodegenerative Diseases (NeuroMMSig): a web server for mechanism enrichment.'

  Bioinformatics (2017): 3679-3681

- Reagon Karki, **Alpha Tom Kodamullil,** Martin Hofmann-Apitius. 'Comorbidity analysis between Alzheimer's disease and Type 2 Diabetes Mellitus based on shared pathways and the role of T2DM drugs'

  Journal of Alzheimer's disease (2017): 721-731

- Juliane Fluck, Sumit Madan, Sam Ansari, **Alpha Tom Kodamullil**, Reagon Karki, Majid Rastegar-Mojarad, Natalie L Catlett, William Hayes, Justyna Szostak, Julia Hoeng, and Manuel Peitsch. 'Training corpora for the extraction of causal relationships coded in Biological Expression Language (BEL)'

  Database (2016):baw113

- Erfan Younesi, Ashutosh Malhotra, Michaela Gündel, Phil Scordis, **Alpha Tom Kodamullil**, Matt Page, Bernd Müller, Stephan Springstubbe, Ullrich Wüllner,

Dieter Scheller and Martin Hofmann-Apitius. ‘PDON: Parkinson’s disease ontology for representation and modeling of the Parkinson’s disease knowledge domain’

Theoretical Biology and Medical Model. (2015):12-20

- Martin Hofmann-Apitius, Gordon Ball, Stefan Gebel, Shweta Bagewadi, Bernard de Bono, Reinhard Schneider, Matt Page, **Alpha Tom Kodamullil**, Erfan Younesi, Christian Ebeling, Luc Canard and Jesper Tegner. ‘Bioinformatics Mining and Modelling Methods for the Identification of Disease Mechanisms in Neurodegenerative Disorders’

  International Journal of Molecular Sciences (2015):29179-206

Contents

Chapter 3

Mapping investment to possible drug target mechanisms

List of Abbreviations

NDDs	Neurodegenerative Diseases
AD	Alzheimer's disease
DALYs	Disability Adjusted Life Years
AICD	Amyloid precursor protein intracellular domain
MAPT	Microtubule Associated Protein Tau
SNP	Single-nucleotide polymorphism
EHRs	Electronic Health Records
KEGG	Kyoto Encyclopedia of Genes and Genomes
SBML	Systems Biology Markup Language
BioPAX	Pathway exchange language for Biological pathway data
APP	Amyloid Precursor Protein
BACE	Beta-site amyloid precursor protein cleaving enzyme
BEL	Biological Expression Language
HGNC	HUGO Gene Nomenclature Committee
RCR	Reverse Causal Reasoning
NPA	Network Perturbation Amplitude
T2DM	Type 2 Diabetes-Mellitus
R&D	Research and development

Introduction

Advancements in high-throughput technologies and diagnostic techniques have produced big data in the biomedical field. However, the question remains whether we are able to retrieve meaningful insights from this vast amount of data to understand disease mechanisms. As we have a limited number of approved drugs for most complex diseases, it remains a challenge to interpret existing data and associate them with molecular mechanisms underlying the disease. The 'one target-one disease' concept adhered to by many pharmaceutical companies, naturally, does not equate to success with regard to complex and multimodal diseases (Achenbach 2011). Accordingly, in order to study complex diseases, it is important to explore approaches that consider whole disease etiology, as opposed to specific perturbed pathways or differential genes. Systems biology is a new avenue to study complex systems and their biological interactions using a holistic approach (Kesic 2016). Systems biology approaches integrate different types of biological data into networks based on interactions among entities (Chen 2013). Biological data can be extremely diverse as it is generated by different groups and is often available in different formats. Moreover, interactions among entities can also be diverse. Based on the different types of interactions between two entities, an entirely different outcome can result, adding complexity to a biological system. Therefore, it is necessary to integrate all combinations of interactions with various types of entities from different sources to understand how a single perturbation in an interaction can elicit and amplify changes in the whole system. Biological networks help to achieve this goal by acting as a platform to integrate heterogeneous data, various causal and correlational interactions, and associations among biological

entities. Successful modeling of diseases necessitates the need for large-scale data-collection and storage, interoperable representation, and development of algorithms and tools enabling pattern or network analysis. This can further help to generate hypotheses to be validated in the laboratory setting, ultimately giving new insights for novel therapeutic targets in the drug discovery processes.

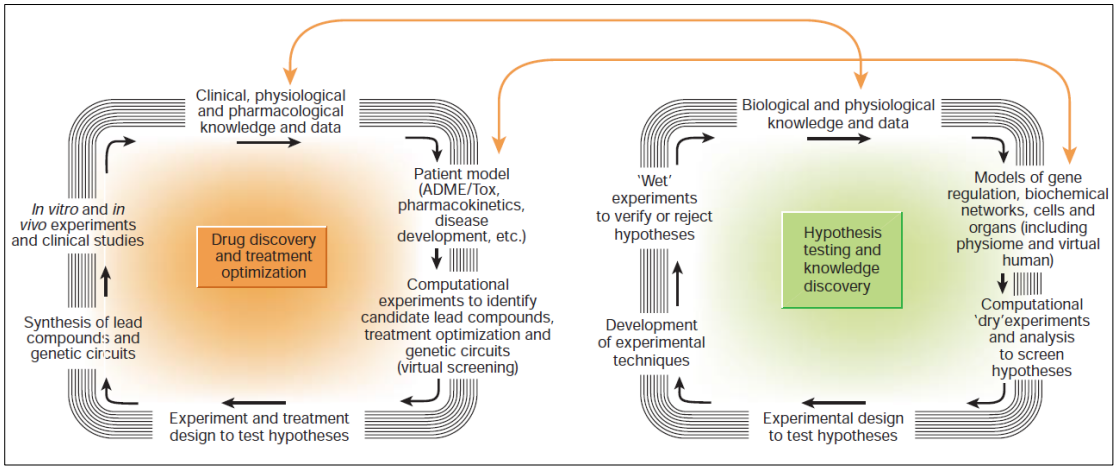


Figure 1: **Interconnection between systems biology and drug discovery cycles** - An illustration of how systems biology approaches like disease models and computational screening and simulation can lead to hypotheses generation, support drug discovery and treatment optimization [Taken From: Kitano, Hiroaki. "Computational systems biology." Nature 420.6912 (2002): 206]

Figure 1 illustrates the analogous processes of systems biology and drug discovery. As shown in the figure, the systems biology model integrates available knowledge and data which describes gene regulatory mechanisms and how it affects biochemical reactions in a system. This leads to further insights into the mechanisms of disease regarding which parts of a system are perturbed in a disease condition. This process promotes hypotheses generation. Moreover, computational disease models allow for better design of in-vitro and in-vivo experiments that lead to drug discovery based on new hypotheses. The next step consists of collecting data from wet lab experiments and converting them

into biological and physiological knowledge. This give rise to treatment optimization to better prioritize targets, and to characterize off-target and on-target effects.

Based on the availability of experimental tissues or cell lines used for experimentation in wet labs, an understanding of the mechanisms involved in many disease areas like cancer, cardiovascular diseases, and infectious diseases have been vastly improved, resulting in greater advancements in therapeutics. However, progress in unraveling biological mechanisms in neurological disorders or neurodegenerative diseases (NDDs), has been limited, in part, to the limited accessibility of experimental tissues and to the complexity of the brain. Additionally, the complexity of the brain is reflected in the heterogeneity of neurodegenerative diseases, which render single drug target approaches ineffective.

Alzheimer's disease and hypotheses on etiology

Amongst the NDDs, Alzheimer's disease (AD) is most prevalent (Figure 2). AD is a progressive neurodegenerative disorder that causes brain cell death. As a result, overall brain size shrinks and the number of connections between neurons decreases, which gradually leads to impairments in memory and cognition (Peters 2006), (Anderton 2002). In the more than one hundred years since Alois Alzheimer identified Alzheimer's disease in 1901, considerable research has been directed towards discovering the cause of the disease (Hippius 2003).

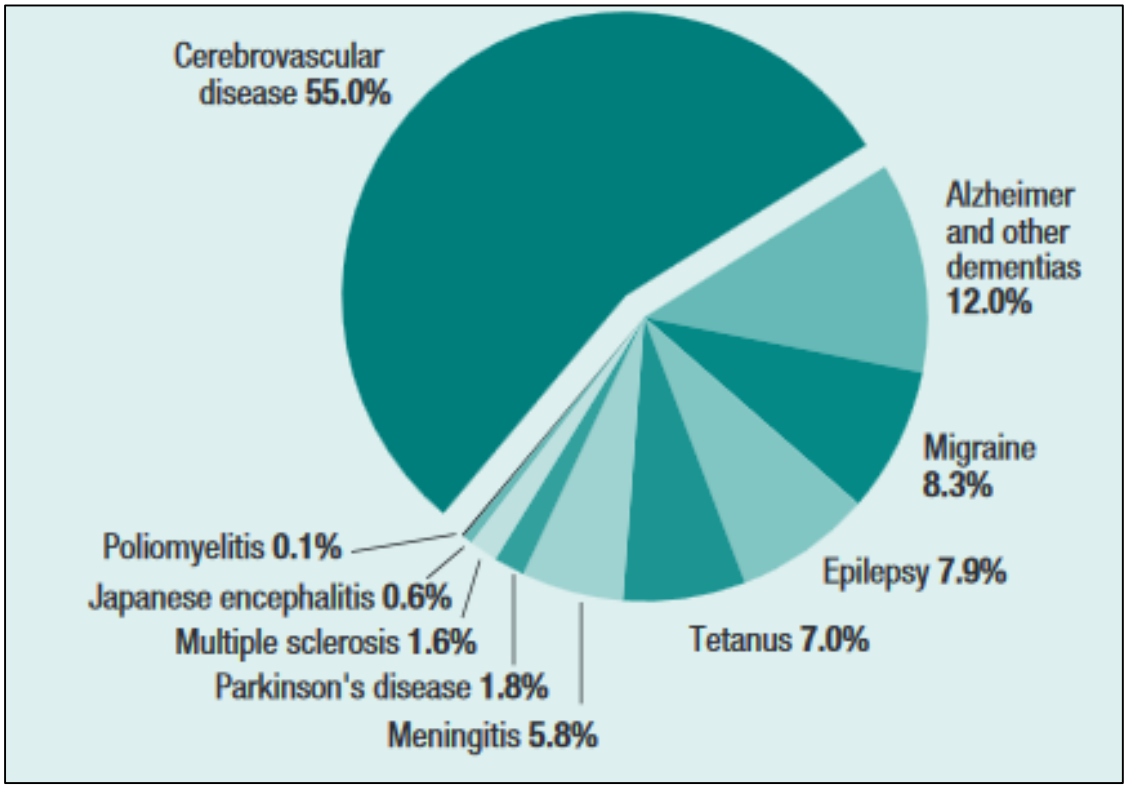


Figure 2: **Percentage of AD among other neurological disorders**–Comparison of Alzheimer's disease with Percentage of DALYs for neurological disorders [Taken from: Global burden of neurological disorders estimates and projections. http://www.who.int/mental_health/neurology/chapter_2_neuro_disorders_public_h_challenges.pdf?ua=1]

Alzheimer's disease is a complex, multifactorial disease, which is driven by biological pathway dysfunction. There are different hypotheses and reasoning about the cause of disease. The major hypotheses under study in AD are outlined below.

Amyloid Hypothesis:

The Amyloid Hypothesis is the most prominent as proposed by the majority of biomedical researchers. It postulates that abnormal proteolytic processing of the amyloid precursor protein (APP) results in the production and aggregation of neurotoxic forms of amyloid ß (Selkoe 2016). APP, a transmembrane protein, is cleaved by alpha-

secretase and beta-secretase, forming soluble fragments such as sAPP-alpha, sAPP-ß, C83 and AICD in the unperturbed pathway. With regard to the pathogenic mechanism, APP is processed by gamma-secretase and produces Abeta 40-42 peptides (Amyloid beta peptides), which form amyloid plaques (Figure 3). These plaques are then deposited in the brain, causing inflammation and other pathway changes leading to AD symptoms.

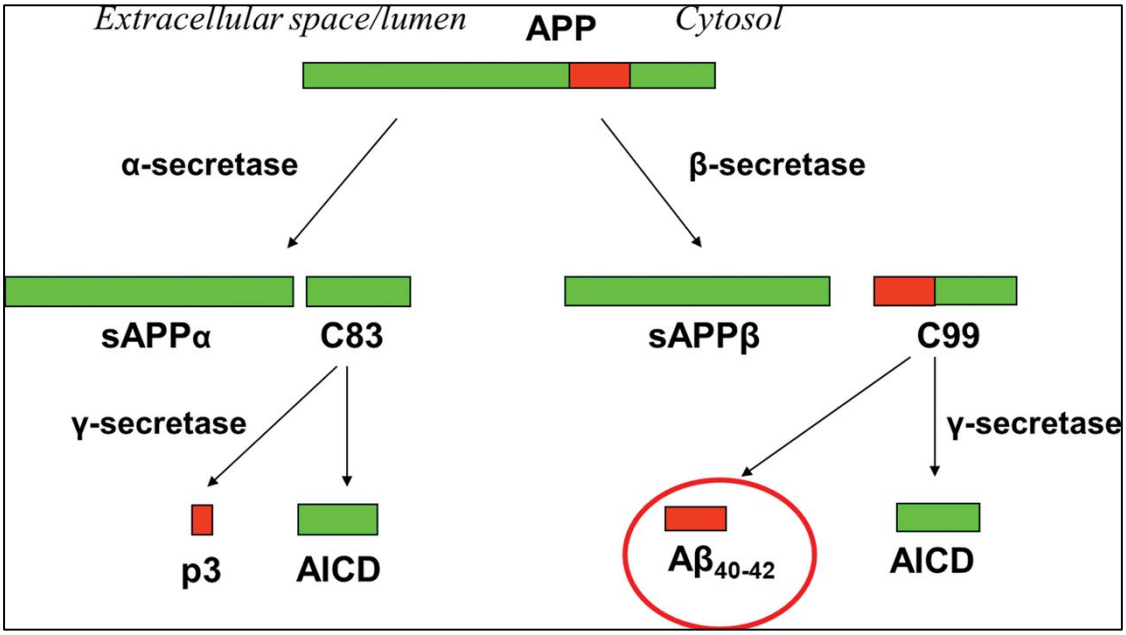


Figure 3: **Processing of APP in healthy brain and in AD brain** - The figure illustrates how processing by different enzymes leads to divergent molecular endpoints. The pathway which lead by alpha-secretase is the normal physiological pathway and the right hand side pathway lead by beta-secretase and gamma-secretase shows the pathological pathway. [Taken From: AF Teich, O Arancio Is the amyloid hypothesis of Alzheimer's disease therapeutically relevant? Biochemical Journal, 2012]

Based on this hypothesis, the following drug target approaches follow to (Alzheimer's Association 2017):

a) Decrease the Abeta 40-42 production by inhibiting gamma-secretase.

b) Inhibit aggregation of and deposition of amyloid plaques

c) Removal of amyloid plaques

d) Reduce inflammation and other symptoms

Recently, the drugs Bapineuzumab, produced by Pfizer and Johnson & Johnson, and Flurizan, developed by Myriad Genetics, both targeting amyloid ß peptides, were withdrawn, as they failed to improve cognitive or functional abilities in patients (Salloway 2014), (Wan 2009). Similar to the above drugs, Semagacestat, a gamma-secretase inhibitor, also failed after being tested in five different clinical trials for not eliciting improvement in cognition. The target, gamma-secretase, also has many physiological functions, such as being a receptor for NOTCH which has a role in neural development (Carroll 2016). APP also plays many vital physiological functions in the brain (Figure 4).

Peptide	Cleavage process	Known function	Reference
sAPPα	α-secretase cleavage of APP	Neuroprotective? Involved in blood clotting?	Smith *et al.* 1990
		Up-regulates BK_{Ca} activity	Furukawa *et al.* 1996
C83	α-secretase cleavage of APP	None known, but may be associated with neurodegeneration	Rockenstein *et al.* 2005
p3	γ-secretase cleavage of C83	None known?	Pardossi-Piquard *et al.* 2005
AICD/CFTγ	γ-secretase cleavage of C83	Transcriptional regulation (e.g. of neprylisin)	Leissring *et al.* 2002
sAPPβ	β-secretase cleavage of APP	Neuroprotective?	
C99	β-secretase cleavage of APP	Altered acetylcholinesterase activity and behavioural influences	Dumont *et al.* 2005
Aβ	γ-secretase cleavage of	Depress synaptic activity	Kamenetz *et al.* 2003
	β-secretase cleaved of APP	Regulation of K_v expression	Ramsden *et al.* 2001; Plant *et al.* 2005
		Regulation of Ca_v	Ramsden *et al.* 2002

Figure 4: **Physiological functions of APP** - Vital physiological functions of APP and processed products in the amyloidogenic pathway [Taken From: Pearson, Hugh A., and Chris Peers. "Physiological roles for amyloid β peptides." The Journal of physiology 575.1 (2006): 5-10.]

The failures of amyloid ß-lowering agents have posed many questions regarding the validity of the amyloid hypothesis. Amyloid ß processing noticeably involves numerous enzymes and signaling pathways, which could play a role in a diverse array of cellular processes. Thus, the clinical failure of amyloid ß -lowering agents does not necessarily amount to the conclusion that the hypothesis itself is incorrect, but rather, it may simply mean that manipulating amyloid ß directly is an unrealistic strategy for therapeutic intervention. However, much research is needed to understand why amyloid beta deposition perturbs the normal signaling pathways in the brain with age.

Tau Hypothesis:

Another notable hypothesis is the Tau hypothesis, which states that the abnormal phosphorylation of MAPT (Microtubule Associated Protein Tau) leads to the formation of hyper-phosphorylated tau and aggregates together to form neurofibrillary tangles in AD (Maccioni 2010). In healthy brains, MAPT organizes microtubule assembly together with tubulin. However, hyper-phosphorylated tau in the diseased brain disintegrates microtubules and promotes formation of neurofibrillary tangles. Deposition of neurofibrillary tangles interferes with axonal transport and leads to cell death (Gong 2008). In 1991, Hardy J et al. reported that it is unclear whether hyper-phosphorylation of tau occurs prior to the formation of neurofibrillary tangles or if it is an outcome of the deposition of amyloid ß peptides (Hardy 1991). As outlined in the amyloid hypothesis, the major drug target approaches in the tau hypothesis consist of reducing tau phosphorylation and neurofibrillary tangles, immunotherapy against tau perturbations, increasing stability of microtubules, and the enhanced removal of neurofibrillary tangles (Godyn 2016). Epothilone D (2013) by Bristol-Myers Squibb, a drug that was recently withdrawn, targeted microtubules to increase stabilization (Alzforum-Epithilone D 2017). Epothilone D showed positive outcomes in mice, but failed in clinical trials (Alzforum-Epithilone D 2017).

Cholinergic Hypothesis:

Apart from the amyloid and tau hypotheses, dysregulation of neuro-chemical signaling pathways involved in AD is a key hypothesis and has been targeted by many drugs is the cholinergic hypothesis (Francis 1999). The cholinergic hypothesis states that the loss of cholinergic function in the brain is associated with the reduced production of the

neurotransmitter acetylcholine (Kihara 2004). The cholinergic hypothesis has gained further prominence as four out of five approved AD drugs target cholinergic mechanisms. In a review, Francis et al. (1999) reported that the impairment of the cholinergic mechanism is evident in the early stages of the disease (Francis 1999). However, some studies suggest that perturbation in the cholinergic mechanism is not the early event in the course of AD, thus challenging the validity of this hypothesis and its potential to be a drug target for disease prevention. For example, Davis KL et al. (1999) reported that the perturbed activity of choline acetyltransferase and acetylcholinesterase were characteristic of a severe AD patient and did not differ significantly from controls in mild AD subjects (Davis 1999). Some recent drugs that targeted the cholinergic mechanism have been discontinued; (ABT-288 (2014) from AbbVie, GSK239512 from GSK (2014), and Varenicline from Pfizer (2011)). ABT-288 aimed to increase the release of neurotransmitters like histamine, acetylcholine, and dopamine, but failed to improve cognitive scores (Alzforum - ABT-288 2017). GSK239512 also had the equivalent mechanism of action as ABT-288, but was withdrawn due to excessive side effects such as headaches, dizziness and a lack of improvement in cognitive and clinical measures (Alzforum-GSK239512 2017). Varenicline is an approved drug to aid in smoking cessation which was in clinical trials for AD. But in the context of AD, it intensifies neuropsychiatric symptoms and gastrointestinal side effects (Alzforum - Varenicline 2017). However, the failure of these drugs do not negate the cholinergic hypothesis, as there are some ongoing trials with positive results such as Encenicline in Phase III, a drug which modulates acetylcholine response in AD (Godyn 2016).

In addition to the aforementioned hypotheses of the underlying pathogenesis of AD, there are many others, including the neuroinflammation hypothesis, oxidative stress hypothesis, glutamate hypothesis, calcium hypothesis, mitochondrial dysfunction hypothesis, cholesterol hypothesis, and metal ion hypothesis (Mohandas 2009).

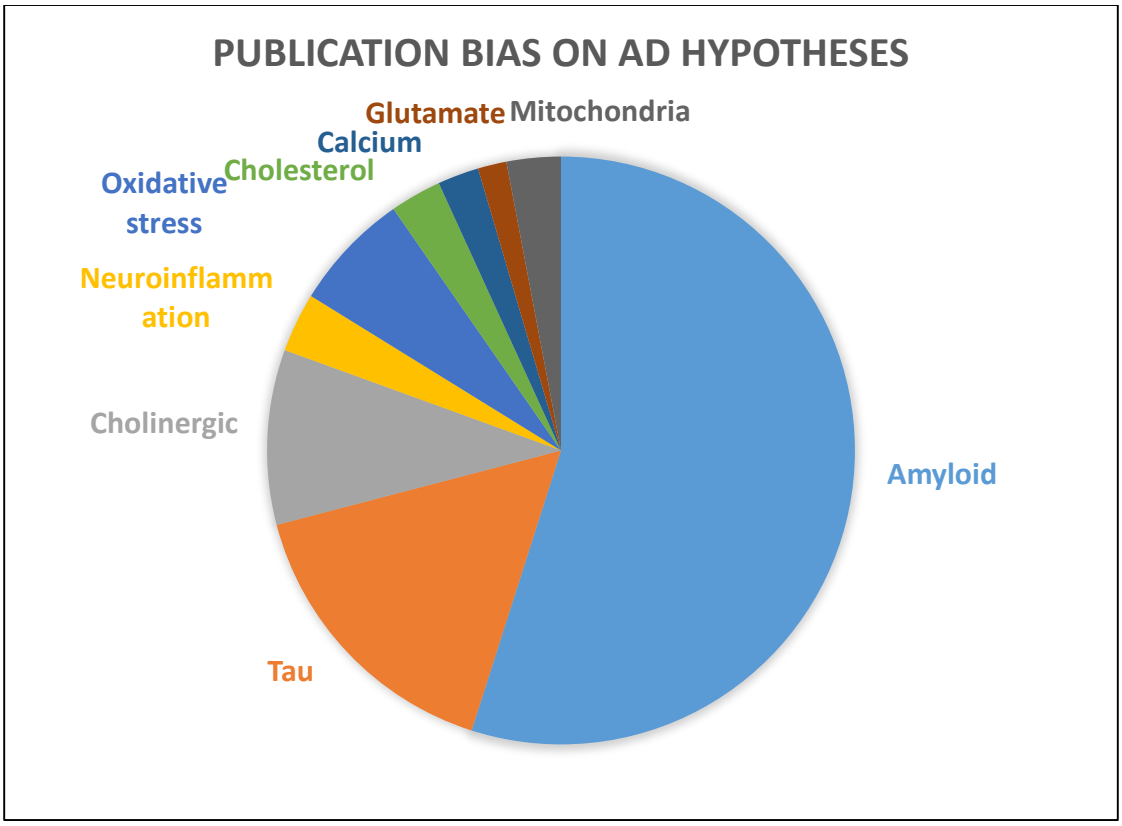


Figure 5: **Comparison of research articles on based on leading disease mechanism hypotheses** related to Alzheimer's disease, from 1997 to 2017, accessed on 15.08.2017 based on PubMed search

Figure 5, depicts the bias in AD research with regard to various hypotheses; more than half of the research between 1997 to 2017 focused on the amyloid hypothesis. However, none of these readily identify the earliest mechanisms of AD which must be targeted to prevent the disease. This necessitates the need for a comprehensive understanding of

the many mechanisms underlying AD along with the normal functioning of the brain, as opposed to a focus limited to specific hypotheses.

Data and knowledge in AD

Knowledge in publications: as text and cartoons

As AD is a multi-factorial disease, discerning disease mechanism in its entirety is possible only through a deep understanding of the interactions between regulatory biological entities (genes, proteins, SNPs, etc) and signaling networks. In order to do so, it is important to assess the current knowledge of the disease, based on signaling pathways. Scientific knowledge about a disease can be accessed from scientific publications, cartoons, and from clinical data. As evident from Figure 6, the research and publications in Alzheimer's disease are increasing over time. However, the assembly of textual knowledge from free-text into biological interaction networks is an immense undertaking, as knowledge from different sources are quite scattered and available in the form of unstructured text.

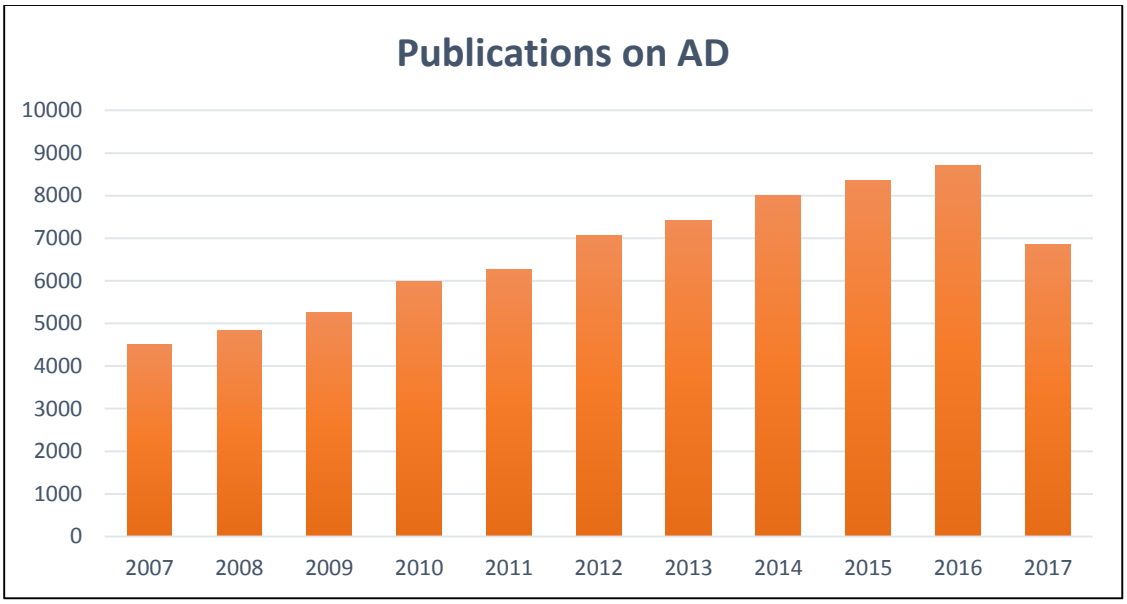


Figure 6: **Estimation of number of PubMed articles on Alzheimer's disease** from the year 2007 – 2017, last accessed on 15.08.2017 from PubMed

Scientific knowledge is diverse and spread over various resources including research articles, reviews, electronic health records (EHRs), patents, and books. In this case, an integrative model that brings this diverse knowledge into a single model is highly advantageous. Unstructured text in publication poses a second challenge. This hinders the extraction of information and the ability to obtain meaningful context without ambiguity. To date, automatic text mining workflows do not exist that can convert these heterogeneous data into a structured format, and subsequently integrate them into networks or models.

In addition to the unstructured and scattered textual knowledge, researchers also represent scientific knowledge as pathway cartoons, which are not computer readable and are difficult to contextualize. For example, figure 7.A shows a pictorial representation of the insulin signaling pathway in a normal, healthy brain whereas Figure 7.B contains more specific information like the role of Aß oligomers in insulin

signaling in an Alzheimer's disease state. It is essential to link together the biological pathways in these figures, as well as capture and store the knowledge and integrate them into a cell signaling network.

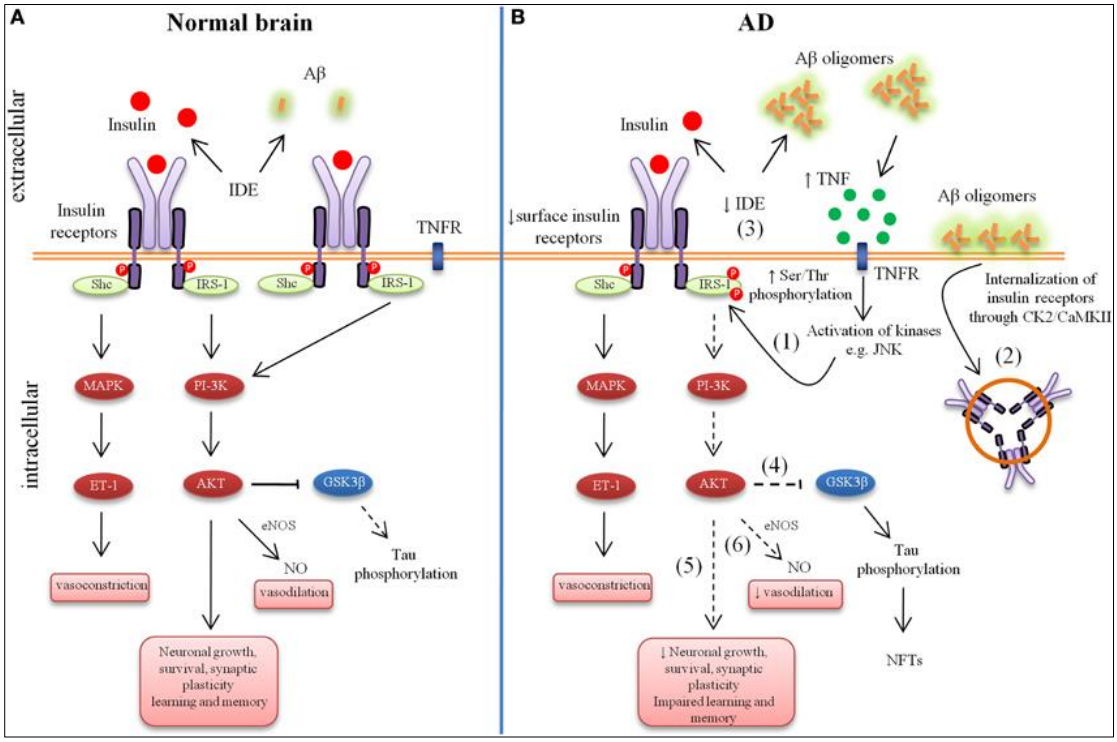


Figure 7: **A pathway cartoon showing normal and disease etiology in Alzheimer's disease based on insulin signaling pathway.** The figure A depicts the normal insulin signaling pathway, while figure B represents the perturbed insulin signaling in AD [Taken From: Bedse, G., Di Domenico, F., Serviddio, G., & Cassano, T. (2015). Aberrant insulin signaling in Alzheimer's disease: current knowledge. Frontiers in neuroscience, 9.]

Apart from the technical difficulties in extracting useful knowledge from text and pictures, reliability of the knowledge itself is questionable. The first issue is publication bias, which can occur due to various reasons. As shown in Figure 4, researchers may focus on well-known hypotheses or established results. Additionally, negative,

insignificant, or novel results may be insufficiently promoted or published, further leading to publication bias (Easterbrook 1991). Nissen et al. (2016) reported that publication bias affects downstream results and demonstrates that positive hypotheses may be canonized as facts much sooner than otherwise (Nissen 2016). It is important to be aware of publication bias as it can affect the results of the integrative approaches. The second issue lies with limited access to clinical research data. Many large pharmaceutical companies or investigators in pharma companies do not publish novel results or disclose data.

One approach to increase reliability of literature-based models is through the use of comparative models. Publications tend to focus on pathological pathways more than normal, healthy function. In complex diseases like AD, it is important to concurrently explore healthy brains as well as various states of disease so comparisons between both are possible. This highlights the major players that perturb the normal functioning of the brain as well as elucidates how AD advances from mild-cognitive impairment to more severe stages of dementia.

With regard to integration of knowledge from pathway cartoons and publications, there are remarkable efforts to store knowledge as pictorial representations. Pathway cartoons of cell signaling pathways are stored in databases such as KEGG (KEGG 2017), REACTOME (REACTOME 2017), and WikiPathways (Wikipathways 2017). These pathway databases serve as a platform to collect various cell signaling pathways which have been established through intervention studies. These databases provide pathways in standard formats like Systems Biology Markup Language (SBML) and Biological Pathway Exchange (BioPaX), enabling easy exchange of data and implementation into

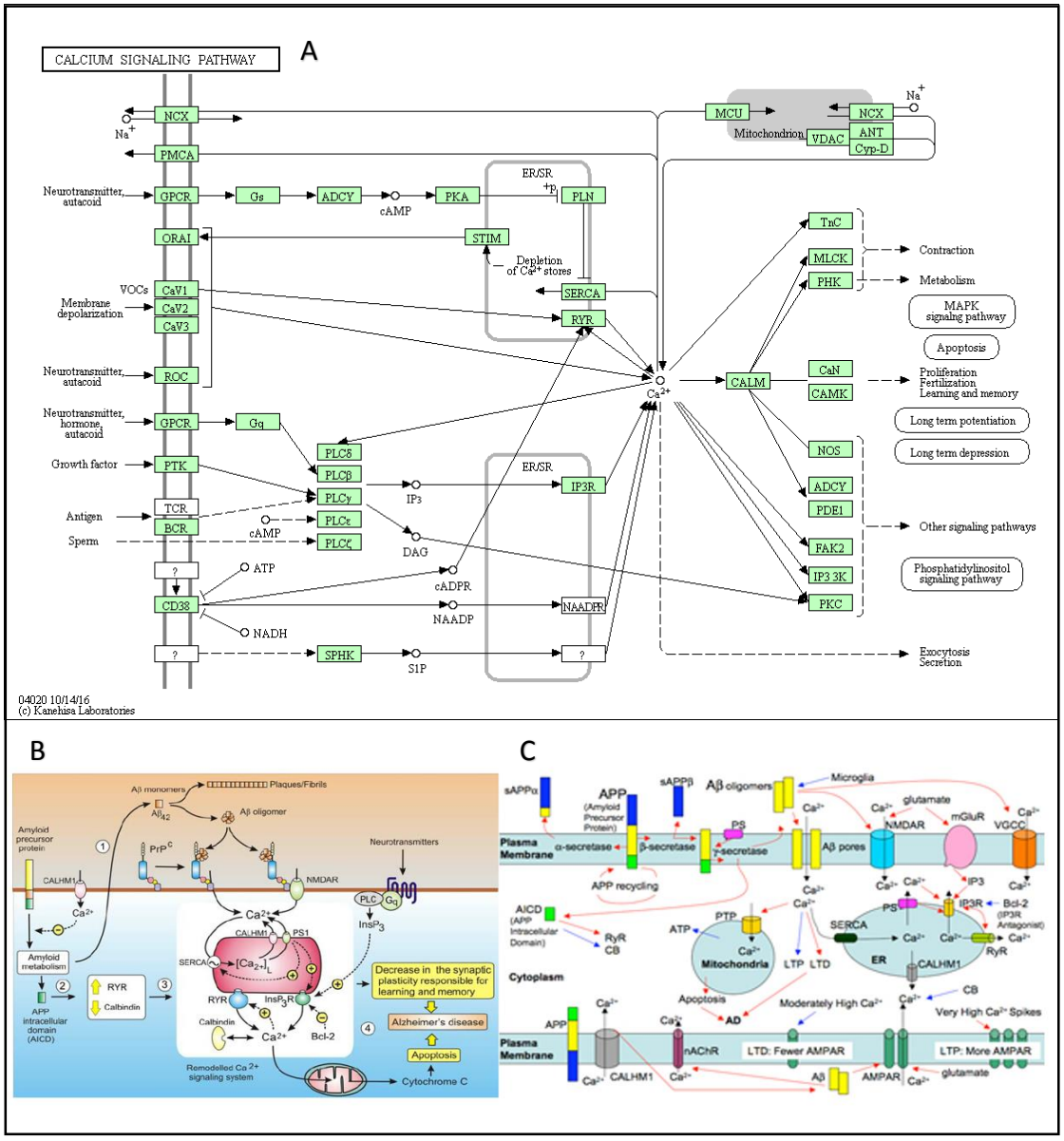


Figure 8: **Comparison of KEGG and Cartoons on the basis of calcium signaling pathway –** The first figure A, depicts the Kegg cartoon of calcium signaling pathway. Whereas figure B and C shows calcium signaling pathway cartoons from a scientific article and from a research site. Each figure contains different information even though all three figures represent calcium signaling.
[Taken From:
1) http://www.genome.jp/kegg-bin/show_pathway?hsa04020+5156+5159
2) Berridge, M. J. (2010). Calcium hypothesis of Alzheimer's disease. PflügersArchiv-European Journal of Physiology, 459(3), 441-449.
3) http://physics.usf.edu/faculty/gullah/]

algorithms for visualization and analysis (Bauer-Mehren 2009). Databases such as

REACTOME contain manually curated data, which are species specific and have cross

references to other established publically available databases (Bauer-Mehren 2009). Conversely, KEGG and WikiPathways are not species or tissue specific and, in the case of Wikipathways, curation is done by crowdsourcing.

With regard to the representation of disease specific knowledge, canonical pathways present some limitations. The first drawback is that pathway databases are not specific to any particular disease, and therefore, corresponding molecular interactions are not disease or tissue specific. Second, pathway databases suffer from inadequate amounts of data and knowledge. This refers to the fact that many of the disease specific canonical pathway cartoons are incomplete, static over time, and not updated regularly as new knowledge is revealed. Third, these pathways are not readily computer readable. Figure 8, depicts a calcium signaling pathway from KEGG and two pathway cartoons from scientific article and website regarding calcium signaling in AD. In the pathway cartoon, specific proteolytic fragments of APP like Aß42, AICD (amyloid precursor protein intracellular domain), and entities which play crucial roles in the amyloidogenesis process, such as BACE and gamma-secretase, are represented. These specific entities are not included in the general KEGG calcium signaling pathway (KEGG, Calcium Signalling Pathway 2017). Furthermore, from the corpus needed to build the KEGG calcium signaling pathway, it is clear that the pathway was based on many different diseases (evident from the annotations to 23 diseases). Therefore, in order to study disease-specific mechanisms, it is necessary to represent and integrate all of the knowledge dispersed as cartoons in the literature and analyze the data together. Until now, there was no automatic way to convert textual knowledge and pathway cartoons into a computer readable form and link this to other biological data. Integration of

knowledge from cartoons as well as unstructured text present in scientific articles is necessary in the age of big data. Hence, developing network models that integrate all interactions between regulatory genes and signaling networks driving the neuropathological pathway of Alzheimer's disease are essential.

Integration of multi-scale data

Apart from the challenges mentioned above regarding pathway cartoons and textual knowledge, integrating multi-scale biological entities and their interactions is a significant hurdle. Moreover, diversified formats of data make linking of data and knowledge tedious. Biomedical data comes from many levels, e.g. genomic, molecular, cellular, clinical, and phenotypic (Figure 9). At the genomic level, data varies from the complete deep sequencing of the genome to narrow information about genetic variants, epigenetic modifications, and differential expression of genes. At the molecular level, data may be various protein interactions from proteomics experiments or metabolomics changes, measured at the clinical level. Various imaging diagnostic techniques generate large amounts of data related to brain function, localized properties of tissue density, local diffusion properties, structural connections between regions, and electrophysiological measures of neural activity. As we acquire more data and knowledge around diseases, the complexity of data accordingly rises. Therefore, integration tools and modeling languages should be capable of warehousing diverse experimental data and knowledge (that are dispersed, different, and difficult to process) into one platform for further analysis.

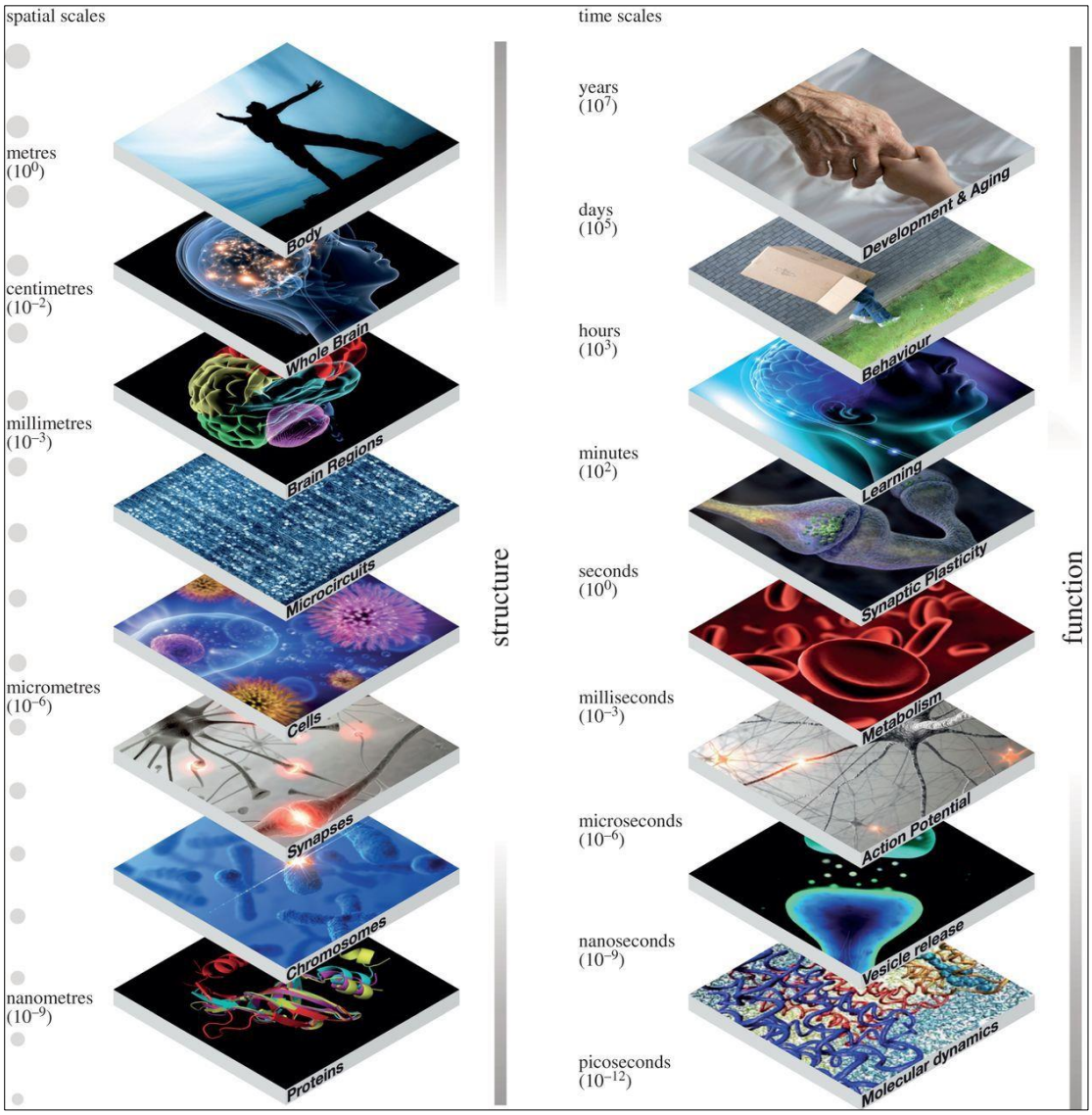


Figure 9: **Multi-scale of biology in the brain**: Different levels of brain physiology based on structural and functional aspects.
[Taken From: http://bouchardlab.lbl.gov/research-and-developement/ Last accessed on March 2017]

Drug discovery in AD

As outlined in the previous sections, research in AD has generated an enormous quantity of data and knowledge. However, this explosion of data has not led to a wealth of new treatments. Currently, there are only 5 approved drugs in AD: Donepezil (1996),

Rivastigmine (2000), Galantamine (2001), Memantine (2003) and a combination drug of memantine and donepezil (2014). These approved drugs do not halt the progression of disease or cure the disease, but rather reduce some symptoms. Until now, more than 800 clinical trials have been conducted new treatments. Figure 10 depicts the number of clinical trials conducted in the past 10 years in AD for various purposes.

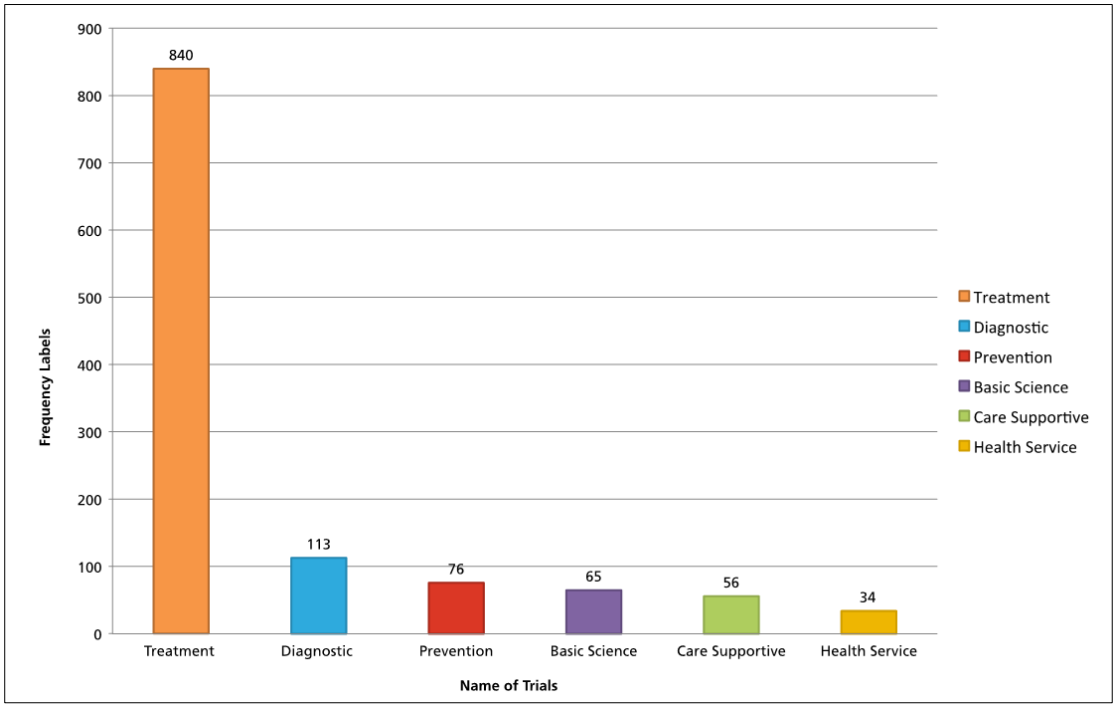


Figure 10: **Number of clinical trials in the last 10 years based on the different types of clinical trials**. In total there were 1550 AD clinical trials of which 1186 AD trials were aimed at treatment, prevention and diagnosis and as supportive trials [Last accessed on March 2016 from https://clinicaltrials.gov/]

Jeffrey L Cummings et. al (2014) and Paul et al (2010) showed that the overall success rate of AD clinical trials is less than 0.4% (Figure 11) (Cummings 2014) (Paul 2010). Many clinical trials are discontinued after Phase II or III due to the following:

1. They fail to demonstrate performance better than placebo

2. Inefficacy of drugs to induce desirable effects

3. Severe side effects of drug usage

4. Lack of efficiency

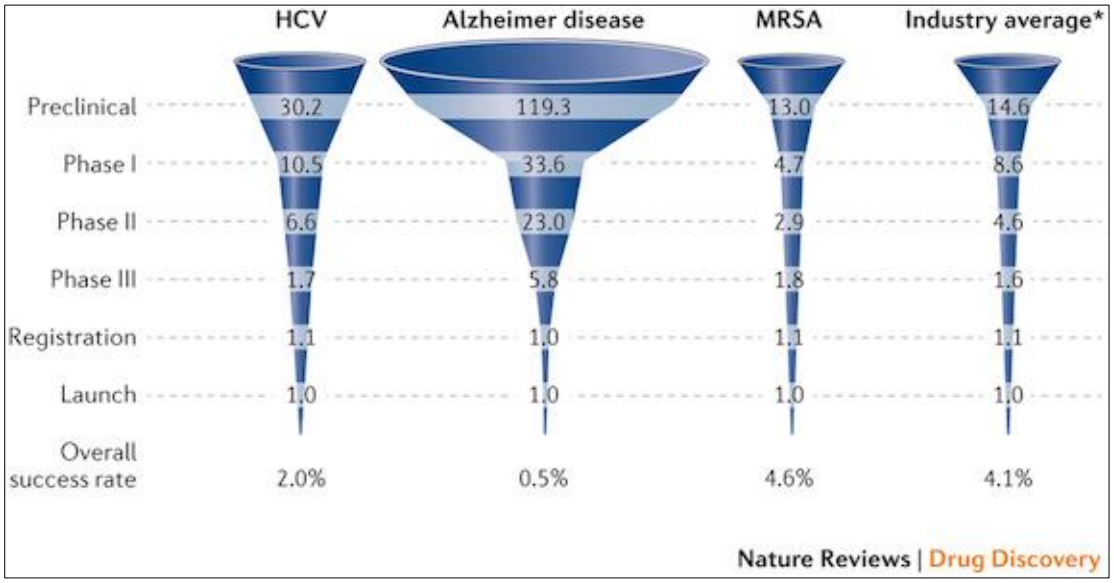


Figure 11: **Overall success rate of AD drugs in comparison to other diseases areas and overall industry average** [Taken From: Calcoen, D., Elias, L., & Yu, X. (2015). What does it take to produce a breakthrough drug? Nature Reviews Drug Discovery, 14(3), 161-163.]

Overall, 72% of agents fail in Phase I, 92% fail in Phase II, and 98% fail in Phase III of clinical trials (Cummings 2014) (Paul 2010). Although new insights about novel potential targets and mechanisms have been provided in recent studies, the question arises of whether pharma companies are in fact investing in the right mechanisms or potential therapeutic targets. Are pharma companies paying enough attention to novel drug-targetable mechanisms and targets?

Reliability of pre-clinical models

Another reason for high attrition rates is due to the fact that pre-clinical models poorly mimic human drug response. Although mouse models contributed significantly to unwinding the etiology of the disease, the translation of genetic responses from mouse to human still remains a complex issue. According to Mouse Genome Informatics, there

are 151 mouse models available for the Alzheimer phenotype (Mouse Genome Informatics - Mouse models 2017). However, no model thus far has contributed to finding a disease modifying therapy in AD (Sabbagh 2013). Additionally, many of these models fail to show the phenotypic endpoints seen in AD patients. As an example, the first animal model of AD was the PDAPP mouse. While this model shows increase in tau phosphorylation, it does not produce neurofibrillary tangles (Sabbagh 2013). One-fifth of mouse coding genes are different from the human genome (Science Daily 2009). When compared on the basis of brain physiology, humans and mice are not analogous models. As depicted in Figure 12, in the mouse brain, prominent regions such as the frontal lobe, parietal lobe, temporal lobe and pons are absent.

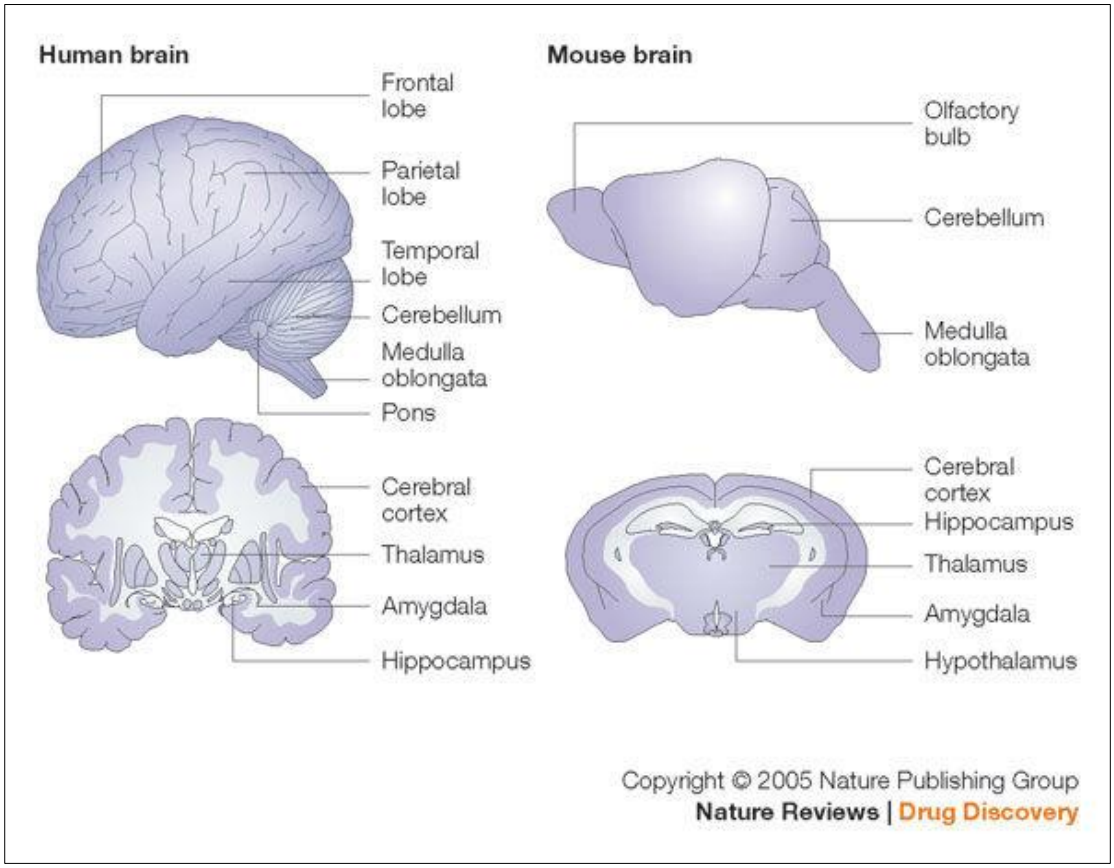


Figure 12: **Physiological difference between mouse and human brain structures** [Taken From: Cryan, J. F., & Holmes, A. (2005). The ascent of mouse: advances in

modeling human depression and anxiety. Nature reviews Drug discovery, 4(9), 775-790.]

Regardless of whether the plethora of animal models suitable for AD research are questionable in translatable, transgenic models of AD continue to gain credibility as more features of the human disease are shown to be successfully represented in mice. As the number of publications and research in AD is increasing, most of this research is done in transgenic models. Thus, the immense need to annotate and use scientific knowledge carefully, as translation of genetic responses among species can vary. As of now, no computable models are available to facilitate the comparability of normal and diseased states or animal and human models and elucidate the cause-and-effect phenomena in molecular cascades and go beyond simplistic cartoon representations of signaling pathways.

In order to fill the gap between research and drug development, we need to bring together state-of-the-art knowledge and data, precisely differentiate translation ability between species and invest wisely in drug-targetable mechanisms. Due to these reasons, it is advantageous to have a knowledge based model to:

(1) aggregate all relevant knowledge

(2) integrate upward and downward regulatory circuits at a maximum granular level at different stages of the disease

(3) predict the drug targets which could effectively be applicable in humans.

To integrate the knowledge and data, it is necessary to adopt computational modeling of mechanisms of disease. In this work, the formalism which is used for modeling is called Biological Expression Language (BEL) (Open BEL 2017).

Computable Modeling of Diseases using BEL (Biological Expression Language)

The BEL language is a formalism for representing the knowledge in life sciences in a computable form. BEL is designed to represent scientific findings by capturing causal and correlative relationships in context, where contextual information may encapsulate biological and experimental system characteristics where the relationships were observed, the supporting publications cited, and the process of curation. Knowledge in BEL is expressed as BEL Statements that are stored in BEL documents. BEL documents are structured text documents that contain BEL Statements, along with sufficient additional information to fully describe and process the document.

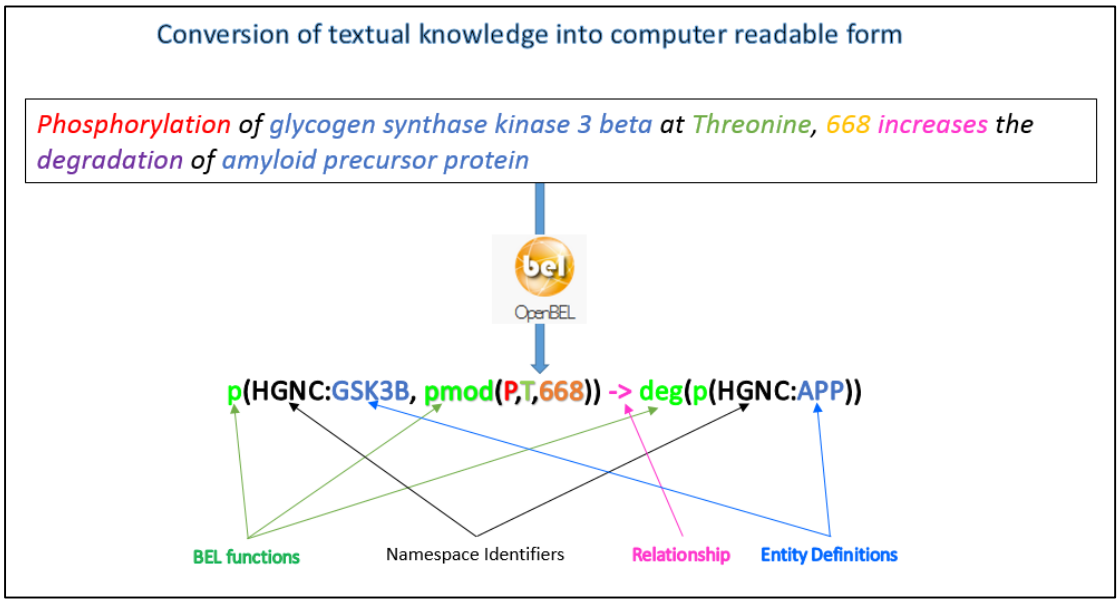


Figure 13: **Example of converting text into BEL –** Illustrated here is the conversion of an evidence from a scientific article. The green highlighted ones are BEL functions, where 'p' indicates *protein*, 'pmod' indicates *protein modification* 'P' indicates *phosphorylation*, 'T' indicates *threonine*, '668' is the *position* of modification. '->' indicates *increase* relationship in BEL and 'deg' function represents *degradation*. All entities like GSK3B and APP is referenced with standard ontology; HGNC.

Apart from a concise set of intrinsic categories of life science concepts, BEL does not prescribe any specific ontology or vocabulary of concepts to be used in the

representation of life science knowledge. Rather, BEL is specifically designed to adopt external vocabularies and ontologies, and therefore, represent life science knowledge in the language and schema of the organization which is collecting or using the knowledge. Thus, biological entities encoded in BEL are defined by reference to values in external vocabularies, which provide a specification of a set of well-known domain values such as the HGNC symbols (Figure 13).

Using BEL, one can encode prior candidate mechanisms of neurodegenerative diseases from publicly available knowledge about a disease with a dedicated syntax ideally suited to model cause-and-effect relationships between biological entities. BEL and the framework are open source and supports algorithms that work on BEL, such as causal reasoning and mechanism enrichment analysis (Catlett 2013). These concepts are already applied in cancer research. One such method is called Reverse Causal Reasoning (RCR), a reverse engineering method to detect mechanistic hypotheses from molecular profiling data. This algorithm gives insights into drug action and toxicity (Catlett 2013). Another established algorithm based on BEL models is the Network Perturbation Amplitude (NPA) which combines high-throughput data and literature derived knowledge to characterize perturbation caused in a collection of biological processes. This framework includes a comparative assessment of the biological impact caused by environmental factors, toxic substances, or drug treatments (Martin 2012). Apart from BEL, there are many other biological expression languages available, such as SBML and Biopathways Exchange Language (BioPAX). Numerous comprehensive disease signaling networks are available with BioPax and SBML in neurology such as

AlzPathway (Ogishima 2016) for AD and molecular interaction map of Parkinson's disease (Fujita 2014).

Even though BEL and BioPAX are both open standards used to capture knowledge about molecular biology and biological processes, their goals differ and they are designed for different applications BioPAX focuses on enabling integration, exchange, visualization, and analysis of biological pathway data. BEL, in contrast, is designed to represent discrete scientific findings and their relevant contextual information as qualitative causal relationships that can drive knowledge-based analytics. In BEL, the focus is on the representation of qualitative causal relationships, capturing statements of cause and effect that enable biological inference by applications. BEL's design enables the representation of causal relationships across a wide range of mechanistic details and between the levels of molecular events, cellular processes, and organism-scale phenotypes. In contrast, BioPAX provides a vocabulary to express control at a precise, biochemical level of description, facilitating the communication of detailed pathway knowledge (Figure 14) (Open BEL 2017).

Figure 14: **Comparison of BioPAX and BEL** for the same evidence text. The same reaction is illustrated in two different formats. [Taken from: https://wiki.openbel.org/]

Until now, there are no fully automated systems to convert knowledge from text into BEL code. However, there are semi-automatic machines available which partially extract BEL statements, though they often require additional manual curation to check the correctness of these statements. The first tool of this type is BelSmile, which is based on a semantic role labelling approach (Lai 2016). Additional tools available in the community are BELMiner (Ravikumar 2017), a rule based relation extraction system, and BELIEF (Fluck 2014) which embeds an information extraction workflow

with state-of-the-art named entity recognition (NER) and relation extraction (RE) methods.

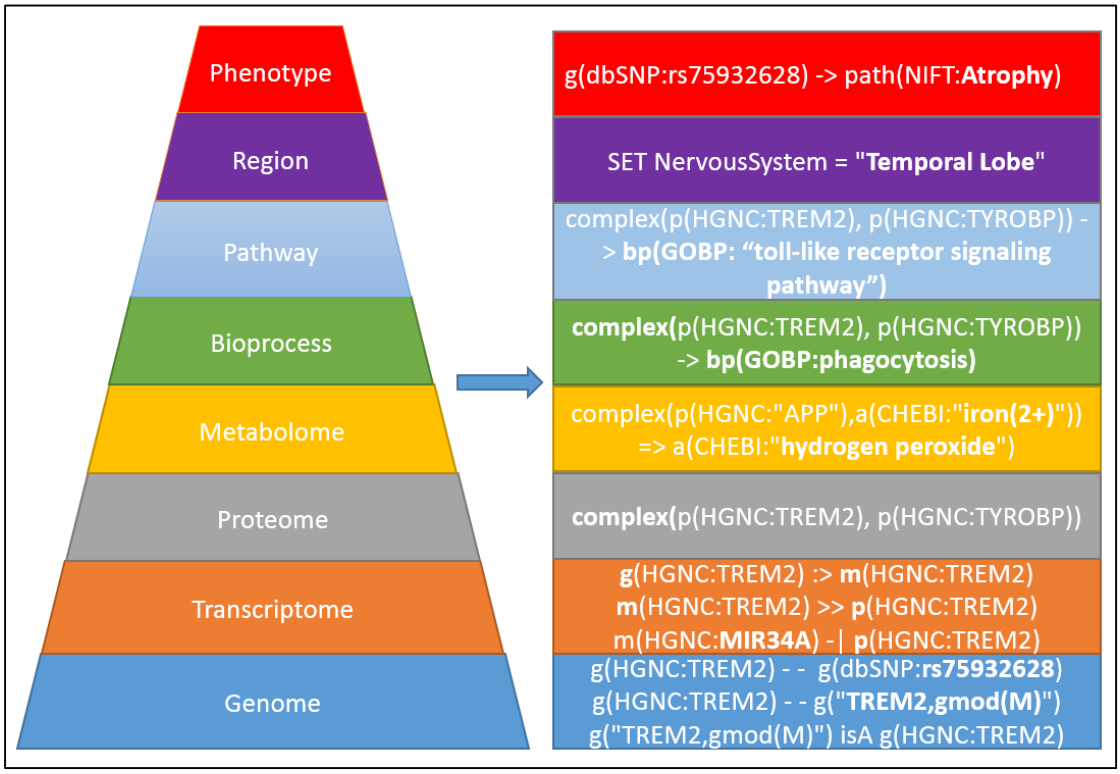


Figure 15: **Examples of representation of different scales in BEL**: this figure illustrates the various functions, relationships and annotations in BEL which are used to represent different levels of biological knowledge from genomic level to phenotypic level.

We have used BEL as the modeling language to build a computational model of AD, which can integrate multiple levels of biological knowledge in a high granularity level. Figure 15 illustrates the integration of heterogeneous, multiscale, and multimodal information in the field of neurology, generally, and neurodegeneration, in particular (Figure 15). We demonstrate, how a combination of text analytics and information extraction with expert knowledge and the abstraction of large-scale experimental data across multiple scales (ranging from omics level to complex clinical readouts) work

together to establish computable "disease models" that are rich sources for new mechanistic hypotheses about the etiology of neurodegenerative diseases.

References

Achenbach 2011: Achenbach, Janosch and Proschak, Ewgenij. 2011. „Rational, computer-aided design of multi-target ligands.“ *Journal of Cheminformatics* P10.

Alzforum - ABT-288 2017: 2017. „Alzforum - Therapeutics.“ *Alzforum - Therapeutics.* Zugriff am 3. October 2017. http://www.alzforum.org/therapeutics/abt-288.

Alzforum - Varenicline 2017: 2017. „Alzforum - Therapeutics.“ *Alzforum - Therapeutics.* Zugriff am 3. October 2017. http://www.alzforum.org/therapeutics/varenicline.

Alzforum-Epithilone D 2017: 2017. „Alzforum-Therapeutics.“ *Alzforum-Epithilone D.* Zugriff am 3. October 2017. http://www.alzforum.org/therapeutics/epothilone-d.

Alzforum - GSK239512 2017: 2017. „Alzforum-Therapeutics.“ *Alzforum-Therapeutics.* Zugriff am 3. October 2017. http://www.alzforum.org/therapeutics/gsk239512.

Alzheimer's Association 2017: 2017. *Alzheimer's Association.* Zugriff am 3. October 2017. https://www.alz.org/national/documents/topicsheet_betaamyloid.pdf.

Anderton 2002: Anderton, Brian H. 2002. „Ageing of the brain.“ *Mechanisms of ageing and development* 811--817.

Bauer-Mehren 2009: Bauer-Mehren, Anna and Furlong, Laura I and Sanz, Ferran. 2009. „Pathway databases and tools for their exploitation: benefits, current limitations and challenges.“ *Molecular systems biology* 5: 290.

Carroll 2016: Carroll, Courtney M and Li, Yue-Ming. 2016. „Physiological and pathological roles of the γ-secretase complex.“ *Brain research bulletin* 126: 199-206.

Catlett 2013: Catlett, Natalie L and Bargnesi, Anthony J and Ungerer, Stephen and Seagaran, Toby and Ladd, William and Elliston, Keith O and Pratt, Dexter. 2013. „Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data.“ *BMC bioinformatics* 14: 340.

Chen 2013: Chen, Bor-Sen and Wu, Chia-Chou. 2013. „Systems biology as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering.“ *Cells* 635--688.

Cummings 2014: Cummings, Jeffrey L and Morstorf, Travis and Zhong, Kate. 2014. „Alzheimer’s disease drug-development pipeline: few candidates, frequent failures.“ *Alzheimer's research \& therapy* 6: 37.

Davis 1999: Davis, Kenneth L and Mohs, Richard C and Marin, Deborah and Purohit, Dushyant P and Perl, Daniel P and Lantz, Melinda and Austin, Gregory and Haroutunian, Vahram. 1999. „Cholinergic markers in elderly patients with early signs of Alzheimer disease.“ *Jama* 281: 1401-1406.

Easterbrook 1991: Easterbrook, Phillipa J and Gopalan, Ramana and Berlin, JA and Matthews, David R. 1991. „Publication bias in clinical research.“ *The Lancet* 337: 867-872.

Fluck 2014: Fluck, Juliane and Madan, Sumit and Ansari, Sam and others. 2014. „BELIEF—a semiautomatic workflow for BEL network creation.“ *Proc. 6th Int. Symp. Semant. Min. Biomed* 109-113.

Francis 1999: Francis, Paul T and Palmer, Alan M and Snape, Michael and Wilcock, Gordon K. 1999. „The cholinergic hypothesis of Alzheimer’s disease: a review of progress.“ *Journal of Neurology, Neurosurgery \& Psychiatry* 66: 137-147.

Fujita 2014: Fujita, Kazuhiro A and Ostaszewski, Marek and Matsuoka, Yukiko and Ghosh, Samik and Glaab, Enrico and Trefois, Christophe and Crespo, Isaac and Perumal, Thanneer M and Jurkowski, Wiktor and Antony, Paul MA and others. 2014. „Integrating pathways of Parkinson's disease in a molecular interaction map.“ *Molecular neurobiology* 49: 88-102.

Godyn 2016: Godyn, Justyna and Jonczyk, Jakub and Panek, Dawid and Malawska, Barbara. 2016. „Therapeutic strategies for Alzheimer's disease in clinical trials.“ *Pharmacological Reports* 68: 127-138.

Gong 2008: Gong, C-X and Iqbal, K. 2008. „Hyperphosphorylation of microtubule-associated protein tau: a promising therapeutic target for Alzheimer disease.“ *Current medicinal chemistry* 15: 2321-2328.

Hardy 1991: Hardy, John and Allsop, David. 1991. „Amyloid deposition as the central event in the aetiology of Alzheimer's disease.“ *Trends in pharmacological sciences* 12: 383-388.

Hippius 2003: Hippius, Hanns and Neundörfer, Gabriele. 2003. „The discovery of Alzheimer's disease.“ *Dialogues in clinical neuroscience* 101.

KEGG 2017: 2017. *KEGG.* Zugriff am 6. Oct 2017. http://www.genome.jp/kegg/.

KEGG, Calcium Signalling Pathway 2017: 2017. „KEGG, Calcium Signalling Pathway.“ *KEGG, Calcium Signalling Pathway.* Zugriff am 3. October 2017. http://www.genome.jp/dbget-bin/www_bget?pathway+hsa04020.

Kesic 2016: Kesic, Srdjan. 2016. „Systems biology, emergence and antireductionism.“ *Saudi journal of biological sciences* 584-591.

Kihara 2004: Kihara, Takeshi and Shimohama, Shun. 2004. „Alzheimer's disease and acetylcholine receptors.“ *Acta neurobiologiae experimentalis* 64: 99-106.

Lai 2016: Lai, Po-Ting and Lo, Yu-Yan and Huang, Ming-Siang and Hsiao, Yu-Cheng and Tsai, Richard Tzong-Han. 2016. „BelSmile: a biomedical semantic role labeling approach for extracting biological expression language from text.“ *Database* 2016: baw064.

Maccioni 2010: Maccioni, Ricardo B and Far{\'\i}as, Gonzalo and Morales, Inelia and Navarrete, Leonardo. 2010. *Archives of medical research* 41: 226-231.

Martin 2012: Martin, Florian and Thomson, Ty M and Sewer, Alain and Drubin, David A and Mathis, Carole and Weisensee, Dirk and Pratt, Dexter and Hoeng, Julia and Peitsch, Manuel C. 2012. „Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks.“ *BMC systems biology* 6: 54.

Mohandas 2009: Mohandas, E and Rajmohan, V and Raghunath, B. 2009. „Neurobiology of Alzheimer's disease.“ *Indian journal of psychiatry* 51: 55.

Mouse Genome Informatics - Mouse models 2017: 2017. *Mouse Genome Informatics - Mouse models.* Zugriff am 8. July 2017. http://www.informatics.jax.org/allele/summary?phenotype=Alzheimer%0D%0A&nomen=&chromosome=any&cm=&coordinate=&coordUnit=bp.

Nissen 2016: Nissen, Silas Boye and Magidson, Tali and Gross, Kevin and Bergstrom, Carl T. 2016. „Publication bias and the canonization of false facts.“ *Elife* 5: e21451.

Ogishima 2016: Ogishima, Soichi and Mizuno, Satoshi and Kikuchi, Masataka and Miyashita, Akinori and Kuwano, Ryozo and Tanaka, Hiroshi and Nakaya, Jun. 2016. „AlzPathway, an updated map of curated signaling pathways: towards deciphering Alzheimer’s disease pathogenesis.“ *Systems Biology of Alzheimer's Disease* 423-432.

Open BEL 2017: 2017. *Open BEL .* Zugriff am 7. October 2017. http://openbel.org/.

Paul 2010: Paul, Steven M and Mytelka, Daniel S and Dunwiddie, Christopher T and Persinger, Charles C and Munos, Bernard H and Lindborg, Stacy R and Schacht, Aaron L. 2010. „How to improve R\&D productivity: the pharmaceutical industry's grand challenge.“ *Nature reviews Drug discovery* 9: 203-214.

Peters 2006: Peters, R. 2006. „Ageing and the brain.“ *Postgraduate medical journal* 84--88.

Ravikumar 2017: Ravikumar, KE and Rastegar-Mojarad, Majid and Liu, Hongfang. 2017. „BELMiner: adapting a rule-based relation extraction system to extract biological expression language statements from bio-medical literature evidence sentences.“ *Database* 2017: baw156.

REACTOME 2017: 2017. *REACTOME.* Zugriff am 5. October 2017. https://reactome.org/.

Sabbagh 2013: Sabbagh, Jonathan J and Kinney, Jefferson W and Cummings, Jeffrey L. 2013. „Animal systems in the development of treatments for Alzheimer's disease: challenges, methods, and implications.“ *Neurobiology of aging* 34: 169-183.

Salloway 2014: Salloway, Stephen and Sperling, Reisa and Fox, Nick C and Blennow, Kaj and Klunk, William and Raskind, Murray and Sabbagh, Marwan and Honig, Lawrence S and Porsteinsson, Anton P and Ferris, Steven and others. 2014. „Two phase 3 trials

of bapineuzumab in mild-to-moderate Alzheimer's disease.“ *New England Journal of Medicine* 322--333.

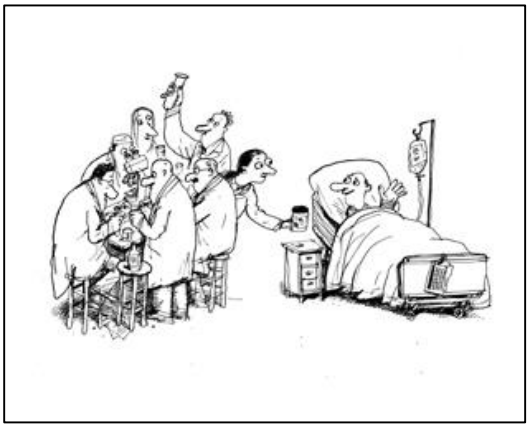
Science Daily 2009: 2009. *Science Daily*. 27. May. Zugriff am 5. October 2017. https://www.sciencedaily.com/releases/2009/05/090526202722.htm.

Selkoe 2016: Selkoe, Dennis J and Hardy, John. 2016. „The amyloid hypothesis of Alzheimer's disease at 25 years.“ *EMBO molecular medicine* 595--608.

Wan 2009: Wan, Hong I and Jacobsen, J Steve and Rutkowski, J Lynn and Feuerstein, Giora Z. 2009. „Translational medicine lessons from flurizan's failure in Alzheimer's disease (AD) trial: implication for future drug discovery and development for AD.“ *Clinical and translational science* 2: 242--247.

Wikipathways 2017: 2017. *Wikipathways.* Zugriff am 5. October 2017. www.wikipathways.org/.

Goals of the thesis



***Bench-to-bedside** – Concept used to describe the process of applying results from biomedical research directly for therapeutic development* [Taken from: https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=561321 https://irp.nih.gov/catalyst/v24i3/news-you-can-use]

The goals of this thesis are outlined below:

- **Develop and implement a strategy to make all relevant knowledge pertinent to Alzheimer's disease available in computable form**

  The first goal of this thesis is to build a high granular multi-scale Alzheimer's disease knowledge model integrating unstructured knowledge. Using this knowledge model I aimed to identify early drug-targetable mechanisms in AD and to understand the functional consequences of genetic and epigenetic variants in disease context.

- **Address reliability of pre-clinical mouse models to mimic human genomic responses**

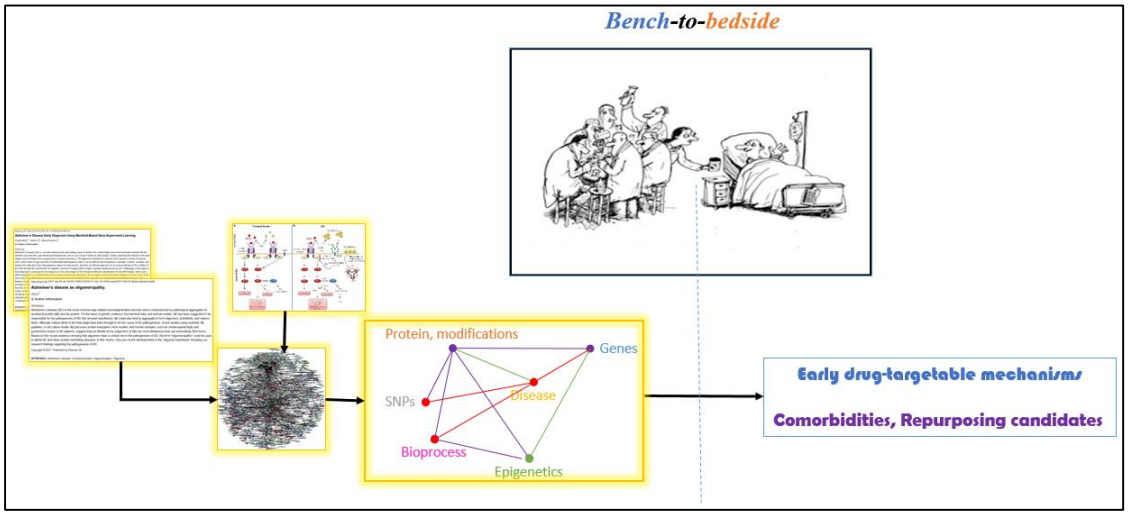
  Having all knowledge around AD in a computable model, how does the comparison with the preclinical models with the human model would look like? Therefore, the second major goal is to assess the degree to which a mouse can translate the genetic responses in human and to what extent mouse model can predict the efficacy of drugs prior to a clinical trial.

- **Establish mechanism inventory and analyze research and development strategy**

  If we have inventory of mechanisms from the established computable cause and effect model, how the investment from pharmaceutical company would appear in the portfolio of mechanisms and do they invest in the right mechanism? Based on this, the third major goal is to investigate the trend of pharma investment on the basis of drug targetable mechanisms in AD.

Chapter 1

Building cause and effect models and predicting novel druggable mechanisms in AD



Introduction

Alzheimer's disease is a complex disorder and the discovery of effective drug candidates to halt the disease requires a deeper understanding of disease mechanisms. This understanding of mechanisms can be made possible by integrating unstructured knowledge and pathway cartoons into a structured computable form which are specific to human. This chapter outlines how an inventory of mechanisms based on cause-and-effect models of Alzheimer's disease and normal, healthy aged brains are built. By comparative analysis, an early stage pathogenic mechanism is identified and new insights are given into shared mechanisms between AD and T2DM based on functional interpretation of genetic variants. This chapter also explains the syntax extension of BEL for the inclusion of epigenetic modifications and demonstrates the relevance of epigenetic modification in NDDs.

Alzheimer's & Dementia 11 (2015) 1329-1339

Featured Article

Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis

Alpha Tom Kodamullil[a,b], Erfan Younesi[a], Mufassra Naz[a,b], Shweta Bagewadi[a,b], Martin Hofmann-Apitius[a,b,*]

[a]*Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany*
[b]*Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn-Aachen International Center for IT, Bonn, Germany*

Abstract

Introduction: The discovery and development of new treatments for Alzheimer's disease (AD) requires a profound mechanistic understanding of the disease. Here, we propose a model-driven approach supporting the systematic identification of putative disease mechanisms.
Methods: We have created a model for AD and a corresponding model for the normal physiology of neurons using biological expression language to systematically model causal and correlative relationships between biomolecules, pathways, and clinical readouts. Through model-model comparison we identify "chains of causal relationships" that lead to new insights into putative disease mechanisms.
Results: Using differential analysis of our models we identified a new mechanism explaining the effect of amyloid-beta on apoptosis via both the neurotrophic tyrosine kinase receptor, type 2 and nerve growth factor receptor branches of the neurotrophin signaling pathway. We also provide the example of a model-guided interpretation of genetic variation data for a comorbidity analysis between AD and type 2 diabetes mellitus.
Discussion: The two computable, literature-based models introduced here provide a powerful framework for the generation and validation of rational, testable hypotheses across disease areas.

Keywords: Alzheimer's disease; OpenBEL; APP; Alzheimer's disease model; Neurotrophin signaling; Type 2 diabetes mellitus

1. Introduction

Difficulties with the diagnosis of Alzheimer's disease (AD) and the absence of disease-modifying treatments for AD remain among the great challenges in biomedicine that need to be addressed in the 21st century. Recent disappointing results of Alzheimer's treatment trials reaffirm that pathogenic mechanisms underlying dementia are more complex than previously thought [1]. Given the obvious complexity of the AD pathology, an important question that arises is whether current knowledge provides a way forward to better understand the underlying pathological pathways.

It has been long hypothesized that the deposition of amyloid-beta peptide in the brain triggers a cascade of molecular events that consequently lead to AD dementia. The amyloid hypothesis represents the mainstream scientific opinion and knowledge on the cause and progression of AD, despite the growing skepticism surrounding this hypothesis [1]. The amyloid-beta protein also plays normal physiologic roles, for example, as protein hormones [2]. Given the amount of accumulated knowledge on both normal and abnormal function of amyloid, which remains scattered in the form of free text and representations in various pathway databases, in silico modeling methods provide a means of aggregating and presenting this information

Competing financial interests: The authors declare no competing financial interests.

*Corresponding author. Tel.: +49-2241-14-2802; Fax: +49-2241-14-2656.
E-mail address: martin.hofmann-apitius@scai.fraunhofer.de

http://dx.doi.org/10.1016/j.jalz.2015.02.006

in a collated, computer-readable format. Major biological processes and pathways involved in the pathogenesis of AD have been collectively represented in the form of the AlzPathway map [3]. However, to be useful and supportive for drug and biomarker discovery efforts, such disease maps need to go beyond the pure representation of pathway information as cartoons, which suffer from missing biological entities (such as Single Nucleotide Polymorphisms [SNPs]) and difficulties in relational representation. Unfortunately, current AD models do not capture the dynamic nature of the disease (e.g. staging) and because of the lack of time course gene expression data on AD in humans, these models do not permit to go beyond the simple overlay of expression snapshots obtained from post-mortem brains.

Future modeling approaches should thus support the automatic reasoning of interlinked molecules and processes. We argue that a computer-processable disease model should be readily amenable to computational reasoning for disease mechanism discovery based on the identification of cause-and-effect regulatory effects, thus linking upstream causal entities to downstream bioclinical effects. Furthermore, in the absence of healthy state models that represent normal cellular processes, any attempt to derive mechanistic interpretations of disease is inconclusive. Thus, disease mechanism discovery requires the conversion of descriptive knowledge into computer-processable cause-and-effect models and mechanistic interpretation should be addressed by the differential analysis of normal and abnormal processes.

We address these requirements by constructing two cause-and-effect computer-processable models for pathophysiological processes associated with AD and their healthy state analogs based on the Biological Expression Language (BEL; http://www.openbel.org/). BEL integrates literature-derived "cause and effect" relationships into network models, which can be subjected to causal analysis using quantitative data such as gene expression. The models developed here not only represent a comprehensive view on the core established pathways involved in amyloid processing but also cover a broad spectrum of events that lead to clinical readouts often seen in AD patients, such as neuroinflammatory processes. Moreover, the healthy and disease state models provide a means for mechanistic differential analysis through which causal pathogenic pathways can be identified.

2. Methods

2.1. Data collection and human APP BEL model building

The scientific knowledge of physiological functions (normal) and pathological actions (diseased) of amyloid precursor protein (APP) processing were acquired from AD-related articles, reviews, and databases. First, 37 pathway cartoons were collected from pathway databases (such as Kyoto Encyclopedia of Genes and Genomes [KEGG] (http://www.genome.jp/kegg/), Reactome (http://www.reactome.org/PathwayBrowser/), and BioCarta (http://www.biocarta.com/genes/index.asp)). Second, using SCAI-View [4] we retrieved a list of 4124 genes, reported to be linked to pathology of AD, of which the top 50 genes were selected based on their relevancy to the query. Documents tagged for these genes were manually filtered for normal (64 documents) and disease (295 documents) states. Relationships reported in these documents were encoded in BEL language v1.0 and used to build the APP BEL models. Furthermore, documents related to top 10 AD related genes were obtained from the AlzGene Database [5]. The APP BEL models were validated for correct syntax and compiled using the OpenBELFramework v2.0, omitting Phase III network augmentation. The models were visualized using Cytoscape and queried using the OpenBEL Knowledge Assembly Model (KAM) Navigator Cytoscape plug-in (https://github.com/OpenBEL/Cytoscape-Plugins).

2.2. Comparison of the normal and disease state models

To identify differential pathways, which are specifically present in the disease state model, the two APP models were compared using the Cytoscape plug-in "advanced network analysis" [6].

2.3. GSEA using MSig database

The "Compute Overlaps" tool available via MSigDB (http://www.broadinstitute.org/gsea/msigdb/help_annotations.jsp#overlap) was used to identify enriched pathways in the BEL models and the AlzPathway map, using the canonical pathways collection of gene sets (MSig database v4.0 updated May 31, 2013). This was used to identify the common canonical pathways between Normal and Disease state BEL models and to revalidate the specificity of models with AD context and to compare it with the existing AlzPathway model. Three canonical pathway data sets were used to compute overlaps; BioCarta, Reactome, and KEGG. Analysis was done using the entire gene list of both (normal and diseased) models and AlzPathway. From the BEL models, we have extracted all the genes/proteins/RNA names (referenced by HUGO Gene Nomenclature Committee [HGNC] namespace) and given as input for computing overlaps. For further analysis, we have selected the top ranked pathways by the number of genes with the highest *P*-value and FDR-q value by Gene Set Enrichment Analysis (GSEA) analysis. From the common pathways, the overlapping genes were identified for both disease and normal BEL models and identified how these pathways differentiate the normal and diseased states.

2.3.1. SNP analysis for comorbidity

Genetic variants (SNPs) for Alzheimer disease (AD) and genes of APP-related pathways were collected from PubMed and genome-wide association studies (GWAS) databases in which SNPs were identified for AD and genes of APP-related pathways. Using GWAS databases,

more than 9000 SNPs for AD with the *P*-value threshold $<10^{-3}$ were collected. Of this, 96 SNPs associated with 47 genes were encoded in the APP-disease model. SNPs were prioritized according to their functional effect on the gene/protein in the disease context based on scores referring to the RegulomeDB database (http://regulome.stanford.edu/index) and experimental evidence for their position in a chromatin state was obtained from the ChroMoS web tool [7].

3. Results

3.1. APP biology models representing normal versus disease processes in human brain

Following the workflow illustrated in Fig. 1, 295 articles were found to contain essential information on APP processing under disease condition in human brain and were used to build the so-called "APP-Disease model"; similarly, 64 articles were used to construct the "APP-Normal model" representing the analogous normal processing of APP in neurons (Fig. 2). This imbalance between the number of articles is reflecting the publication bias toward APP in the disease context as compared with reports on its normal biological role. Although we are aware of this bias, we aimed at the maximum coverage of causal and correlative statements that can be encoded in BEL. Consequently, the models we present here have grown way beyond this well-characterized APP pathophysiological endpoint and now include the vast majority of AD associated processes and pathways. As a result, the models encoded in BEL consist of 701 nodes for "APP-Normal" and 1314 nodes for "APP-Disease". There are 920 BEL knowledge statements configuring the APP-Normal model and 2087 BEL statements supporting the APP-Disease model. The total numbers of interactions (edges) in normal and disease models are 1416 and 2935, respectively.

The APP BEL models (normal and disease) were compared with the previously published AlzPathway model in terms of information coverage [3]. A comparison among all three models is shown in Supplementary Table 1. To investigate functional similarities and differences in content, a comparative pathway analysis was performed with the AlzPathway model using pathway enrichment analysis (see Methods) with the canonical pathways in the MSig database [9] (Supplementary Table 2). The literature supporting the role of unique pathways of APP-Disease model indicates that these unique pathways form the core of hypotheses describing the pathology of AD (Supplementary Table 3)

3.2. Differential analysis of APP-Normal and APP-Disease models for identification of causal events

The differential model analysis aims to identify pathophysiological mechanisms underlying disease in comparison to the normal baseline function. We developed a strategy for differential model analysis that normalizes between the two models at the level of common, overlapping processes and pathways.

After the alignment of two APP models, we identified the disease-specific parts of the APP-Disease model by subtraction of the nodes and edges shared by both models. The resulting "delta" model was subjected to pathway enrichment analysis, which resulted in the identification of several pathways enriched in the portion of the model that is unique to disease, for example, the neurotrophin signaling pathway, mitogen-activated protein kinase (MAPK) signaling pathway, and signaling by nerve growth factor (NGF). The identification of these pathways provides a starting point for the generation of mechanistic hypotheses. The integration of additional information from high-throughput data sources (e.g. GWAS data; gene expression data) and scientific literature not used to build the BEL model (e.g. patent literature) provide independent evidence for the relevance of a putative disease mechanism identified through differential model analysis.

Because the neurotrophin signaling pathway was among the top identified, disease-associated pathways, we investigated this pathway in more detail. In our BEL models, we identified four key regulators of the neurotrophin signaling cascade as described in the KEGG neurotrophin pathway, namely NTRK2 (neurotrophic tyrosine kinase receptor, type 2), BDNF (brain-derived neurotrophic factor), nerve growth factor receptor, and NGF. However, the differential model analysis reveals that the mode of interaction among these four proteins drastically differs between the normal and disease states. Accordingly, these proteins control two branches of the neurotrophin pathway, which regulate the balance between two possible biological outcomes, namely neuron survival versus apoptosis (Fig. 3).

The neurotrophic protein BDNF and its receptor NTRK2 are involved in neuron differentiation and growth. In the normal state, ubiquitin carboxyl-terminal esterase L1 (UCHL1), a deubiquinating enzyme that controls BDNF-mediated retrograde transport, activates BDNF, and increases the binding of BDNF to its receptor NTRK2, thereby promoting neuronal development and homeostasis. In contrast, under AD conditions, amyloid-beta prevents the binding of BDNF to NTRK2 receptor, thereby blocking BDNF-NTRK2 downstream signaling. This blockade leads to the repression of neuron survival, differentiation, and growth, so that abnormal APP processing and amyloid-beta production has been experimentally shown to attenuate BDNF-NTRK2 signaling [10]. UCHL1 activity is repressed by amyloid-beta, which in turn impairs BDNF-NTRK2-mediated downstream signaling, leading to diminished synaptic plasticity and neuronal survival [11]. Our BEL models also shed light on a second pathophysiology mechanism of two other proteins involved in neurotrophin signaling: NGFR and NGF. In the normal state, the NGF protein binds to NGFR resulting in NGFR polyubiquitination. Ubiquitinated NGFR binds to inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta (IKBKB) and activates

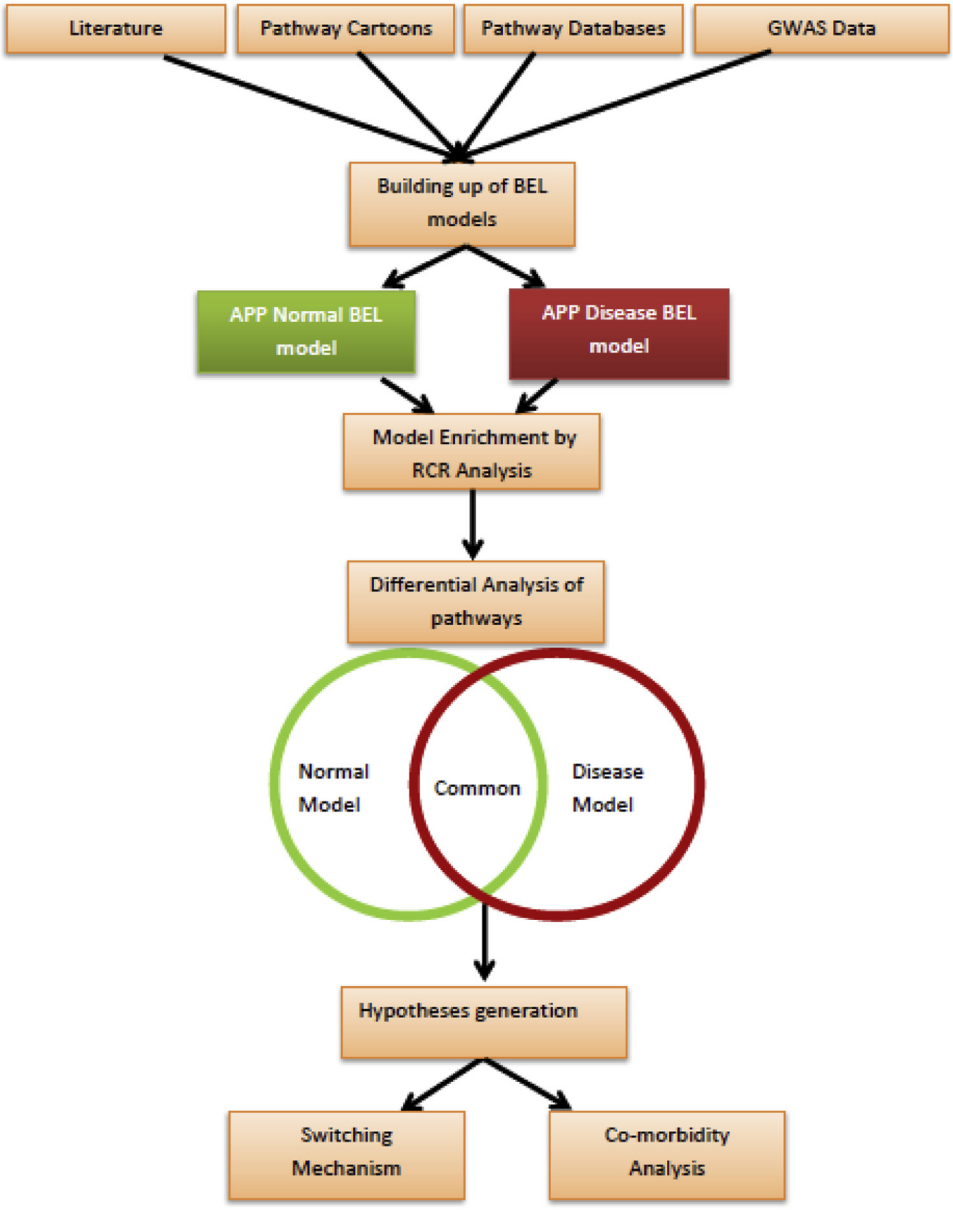


Fig. 1. Schematic representation of the model construction and analysis workflow: Two amyloid precursor protein (APP) models were built using the scientific knowledge present in the scientific literature, databases, pathway cartoons, and genomic databases. The two models represent the normal neuron physiology and the diseased state physiology. The initial models have undergone an enrichment through Reverse Causal Reasoning (RCR) analysis [8]. Differential model comparison based on gene set enrichment led to the generation of two hypotheses, which were investigated further in silico and are supported by additional, independent evidence.

nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1), which promotes neuronal cell survival [11]. In the disease state, amyloid beta peptides competitively bind to the NGFR and inhibit the binding of NGF, resulting in increased cell death [13,14].

An exhaustive search of patent and nonpatent literature for further evidence supporting the mechanism of competitive blocking of the NGF receptor through APP peptides revealed that although the literature supports the inhibition of BDNF signaling by APP and the induction of NGFR-mediated cell death by APP, separately, the embedding of the competitive binding of NGF and APP peptides in the context of the model shown in Fig. 3 brings these observations together as a novel, cohesive disease mechanism. Further supportive evidence comes from the patent literature [15]. Accordingly, the occurrence of a mutation from lysine to alanine at position 34 of the NGF amino acid sequence has been detected that results in binding mutant NGF molecule to NGFR with 50% lower affinity. Interestingly, the patent reports

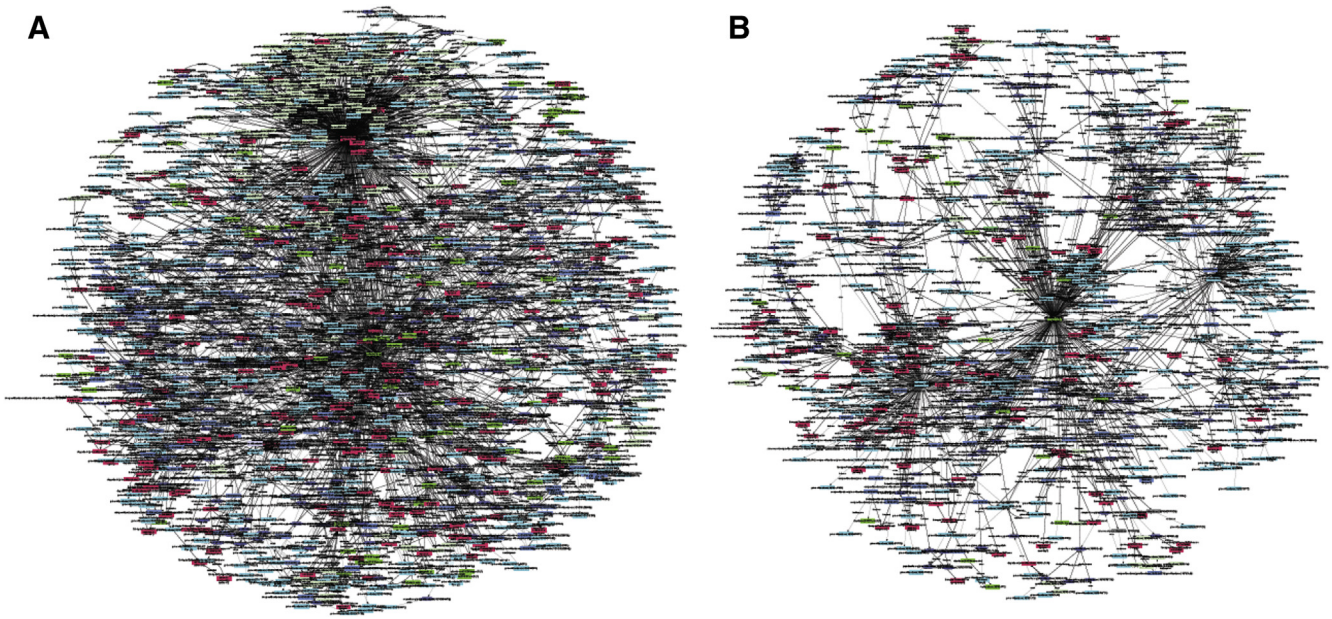


Fig. 2. Amyloid precursor protein (APP) biological expression language (BEL) models: The first image (A) represents APP-Disease model and second image (B) represents APP-Normal model. The observation that the APP-Normal model is sparse but the APP-Disease is denser reflect the bias toward the research in pathology, as compared with normal physiology of APP in the human brain.

a simultaneous occurrence of the amino acid sequence "lysine-glycine-alanine" in the amyloid peptide that provides a binding site for NGFR, thus creating a competitive binding capacity for the amyloid peptide.

3.3. Systematic aggregation of evidence in support of the amyloid-mediated neurotrophin switch hypothesis

The putative amyloid-mediated switch mechanism identified through differential model analysis of the neurotrophin signaling pathway is based on qualitative information. To further support the mechanism of action exerted by amyloid beta in the neurotrophin signaling pathway, we systematically harvested and screened independent pieces of evidence from experimental databases containing data sets on knockout mouse models and miRNAs. For four key regulators in the neurotrophin signaling pathway (BDNF, NGF, NGFR, and NTRK2), knockout mice were identified in the Mouse Genome Informatics database [12] and this provides supportive evidence for the proposed amyloid-switch mechanism (Supplementary Table 4).

We also systematically investigated reports on miRNAs that regulate the genes in the neurotrophin pathway. Indeed, several miRNA studies provide supportive evidence for a key role of members of the neurotrophin pathway in early decision making on neuron survival [16,17] (see Supplementary Table 5).

3.3.1. Biomarker-guided validation of the amyloid-mediated neurotrophin switch hypothesis

Mentions of potential biomarkers in the literature can be used for biomarker-guided pathway analysis [18]. We therefore extracted mentions of potential biomarker functions of BDNF, NTRK2, NGF, and NGFR from the literature (Supplementary Table 6). Mapping these evidences for expressed biomarkers to the neurotrophin pathway clearly supports the amyloid-dependent switch mechanism hypothesis (Fig. 3). The coordinated decrease in the levels of NGF and increased expression of NGFR protein, on one hand, and consistent decrease in levels of BDNF-NTRK2 complex, on the other hand, is aligned with our hypothesis and can be mechanistically explained by the model. In addition, the decreased expression of BDNF and NTRK2 in synergy with the inhibitory effect of amyloid beta on UCHL1 leads to "switching" the entire pathway from its normal state with neuroprotective effect to the disease state with a strong trend toward neuron apoptosis.

3.4. Inclusion of genome variation information in causal models

The addition of information on genetic variation to BEL models can support the generation of new hypotheses and analyze their mechanistic link to comorbidities of AD such as diabetes. The enrichment analysis of our models earlier indicated that insulin signaling pathway is among significantly enriched pathways connected to the APP processing (see Supplementary Table 2). Accumulated evidence suggests that type 2 diabetes mellitus (T2DM) is a strong risk factor for AD, as shown by Akomolafe (2006) [19] and patients treated with insulin were at highest risk of dementia [20].

Consistent with these findings, analysis of APP-Normal and APP-Disease models revealed that some interactions in

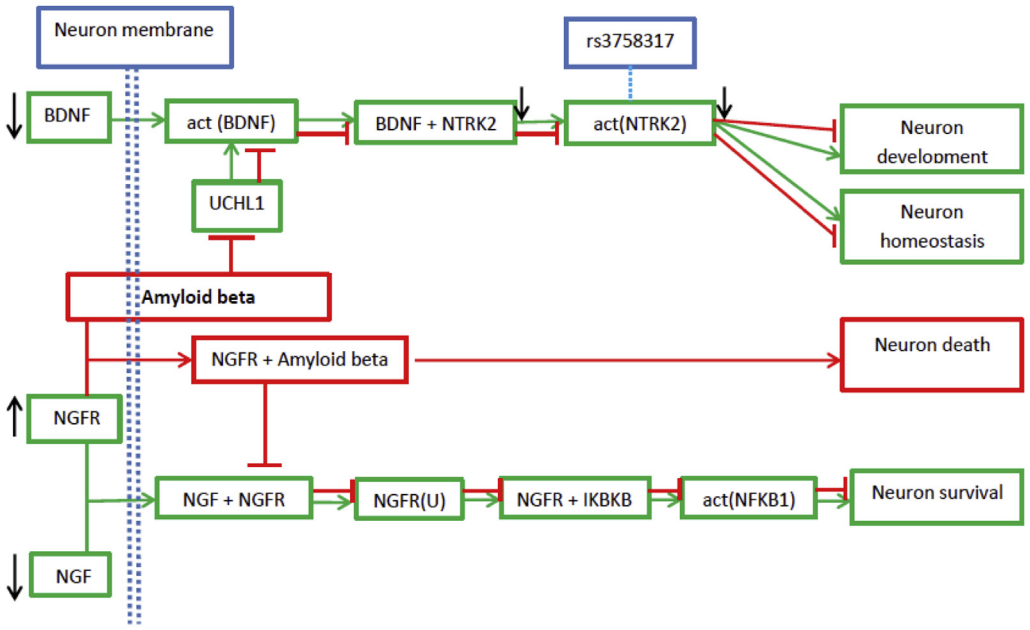


Fig. 3. Molecular decision-making mechanism linked to the neurotrophin signaling pathway between Normal and Disease states: The nodes and edges shown in green color represent the normal pathway. The red color indicates the perturbation of the neurotrophin pathway under Alzheimer's disease diseased state conditions. The node act represents activity of the protein and (U) represents the ubiquitination of the protein. The black colored arrows (up and down) indicate over- or underexpression of the nodes in diseased state from the biomarker guided validation.

normal insulin signaling are perturbed by AD-causing factors such as amyloid beta peptides (see Supplementary Table 2). We have identified 12 functionally relevant SNPs linked to AD and T2DM and associated them with three genes (clusterin [CLU], serine/threonine kinase 11 [STK11], and phosphatidylinositol binding clathrin assembly protein [PICALM]) in the APP-Diseased BEL model (Supplementary Table 7). Two of the three genes (CLU and STK11) could be integrated with prior knowledge to build a hypothesis (Fig. 4).

3.4.1. Model-guided interpretation of genetic variation data by inferring chains of causation

The comorbidity model shown in Fig. 4 reveals the possible causative effects of genetic variants on the mechanistic association of T2DM with AD. In a recent GWAS study on AD, CLU intronic SNPs were found to be associated with the disease [4,21,22]. Clusterin is a transport protein and has a role in helping the clearance of amyloid-beta by transporting it through the blood-brain barrier [23]. The risk variant rs9331888 (with allele G) associated with the CLU gene increases the quantity of a CLU isoform in AD, which induces apoptosis [24] and may contribute to the accumulation of amyloid beta in AD (Fig. 5). A study that quantified levels of clusterin isoforms showed a decrease in secreted soluble CLU in prodromal Alzheimer brain and a significant increase in intracellular CLU [25]. Furthermore, in a proteomic analysis of human hippocampal tissues from AD brains and age-matched control brains, it was confirmed that an isoform of CLU is upregulated in AD cases [26]. Moreover, AD patients have a higher expression of CLU mRNA and its concentration is positively linked to programmed cell death or apoptosis [27]. In line with these observations, expression quantitative trait loci (eQTL) analysis has shown that CLU is overexpressed in AD [28]. It is possibly linked to fibrillar amyloid-beta and apoptotic mechanisms in neurodegenerative diseases [29]. One of the SNPs (rs1532278) in the CLU gene is also associated to T2DM in GWAS analyses of diabetes patients; the amount of CLU is also significantly increased in the serum of T2DM patients, which is correlating with blood glucose levels [29]. It has been hypothesized that this SNP is linked to T2DM through insulin resistance and impairment of insulin secretion.

Similarly, genetic variants of STK11 have been linked to T2DM and also to AD in GWAS studies [39,40]. In mouse models, the deletion of this gene is linked to the inhibition of axon branching [41]. According to GWAS studies, two intronic SNPs in the STK11 gene are associated with AD [39]. Moreover, two intronic SNPs of STK11 are also associated with T2DM. The expression of STK11 in liver seems to be required to lower blood glucose and its deficiency upregulates gluconeogenesis. Additionally, targeted STK11 deletion in liver leads to hyperglycemia [42] (Fig. 6). Furthermore, mouse models lacking S6K1 (C57BL/6J) display enhanced insulin sensitivity [7]. The dysfunction of STK11 gene may also contribute to the accumulation of amyloid beta via the overactivation of mTOR and inhibition of autophagy. Gene expression studies show that STK11 is downregulated in AD [43]. Finally, the relevance of the regulatory and causal

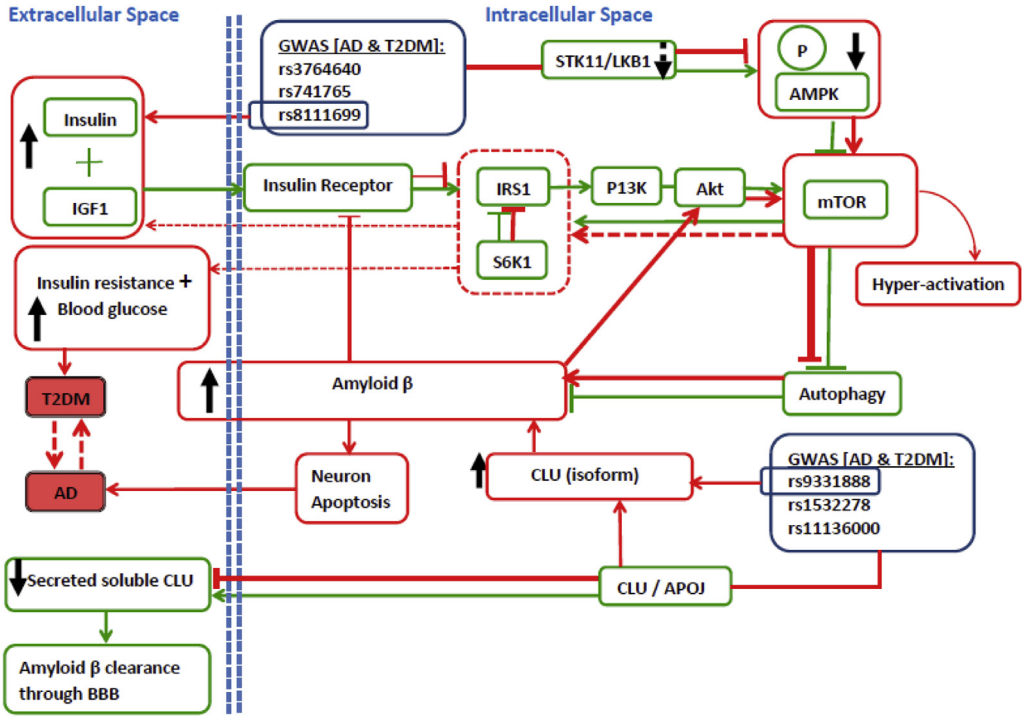


Fig. 4. Comorbidity association of Alzheimer's disease (AD) and type 2 diabetes mellitus (T2DM) by genetic variants of clusterin (CLU) and serine/threonine kinase 11 (STK11) genes: In the normal state (green color edges), insulin protein binds to its receptor insulin receptor and this binding event activates INSR through phosphorylation [30]. The activated INSR binds to insulin-like growth factor 1 (IGF1) and activates insulin receptor substrate 1 (IRS1) [31]. Activated IRS1 activates the phosphoinositol signaling system which activates protein kinase B (AKT) signaling and controls glycogenesis. Activated INSR binding to IGF1 also activates Src homology 2 domain containing protein (SHC) and thereby activates the MAPK signaling pathway [32]. In the disease state (red color edges), CLU promotes neuron apoptosis [27]. Amyloid beta peptides bind to INSR, effectively preventing activation of INSR by insulin. As a consequence, through inactivation of the phosphoinositol signaling system, AKT signaling and mitogen-activated protein kinase (MAPK) signaling pathways, binding of APP peptides suppresses the insulin signaling pathway [33]. The CLU single nucleotide polymorphisms (SNPs) are associated with an increased production of amyloid beta peptides and the CLU variants increase the risk of T2DM by primarily inducing the insulin resistance and secondly by decreasing the production of insulin [34]. In the case of insulin resistance, the amount of INS is increased due to its accumulation in the blood [35]. Normally under the condition of energy stress, STK11 activates adenosine monophosphate-activated protein kinase (AMPK) by phosphorylation and AMPK activation decreases Mechanistic target of rapamycin serine/threonine kinase (mTOR) signaling activity, thereby helping degradation of β-amyloid. In T2DM, the SNP rs8111699, which maps to the enhancer region of the STK11 gene, is influencing insulin sensitivity [35]. The other SNP (rs741765) is located in the insulator region, which may block the interaction between the enhancer and promoter of the gene, resulting in downregulation of the STK11 gene [36]. Deficiency and dysfunction of STK11 inhibits the AMPK phosphorylation, thereby reducing the activity of AMPK [37], which hyper-activates mTOR signaling in AD [38]. Moreover, in T2DM, hyperactivation of mTOR signaling inhibits IRS1 via activation of S6K1 and the IRS1 inhibition leads to insulin resistance (linking the STK11 causal graph to the CLU graph), which leads to increase in INS and glucose in blood. The black colored arrows (up and down) indicate over- or underexpression of the nodes in diseased state; while dotted arrows are inferring the possible effect of genetic variants.

circuitry outlined here is underpinned by the activity of an antidiabetic drug known as metformin, which is used to activate AMPK (Adenosine monophosphate-activated protein kinase) phosphorylation and probably may repress and delay the appearance of AD pathology [44].

4. Discussion

There is an unmet need for strategies to model and identify potential disease-initiating events/mechanisms in the absence of both sufficient data (which makes data-driven approaches impossible) and models for early Neurodegenerative disease (NDD) initiation (which makes a simple cause-effect analysis very difficult). We believe that complex, idiopathic diseases cannot be addressed by the established routes of molecular biology experimentation alone, as neurodegeneration works in the context of an entire organ and the pathology can only be studied in the organ context. Model-driven approaches are a way to capture the collective knowledge about disease processes and allow for a comparison at systems level.

The results of this study demonstrate that encoding relevant knowledge into causal relationship models confers enhanced interpretation power that is well-suited for hypothesis generation. BEL models of APP processing represent a broad coverage of the molecular knowledge on the pathological events underlying AD while preserving sensitivity (by inclusion of various biological pathways linked to the core pathology), specificity (by inclusion of species- and disease-specific information), and context (by inclusion of almost all types of biological

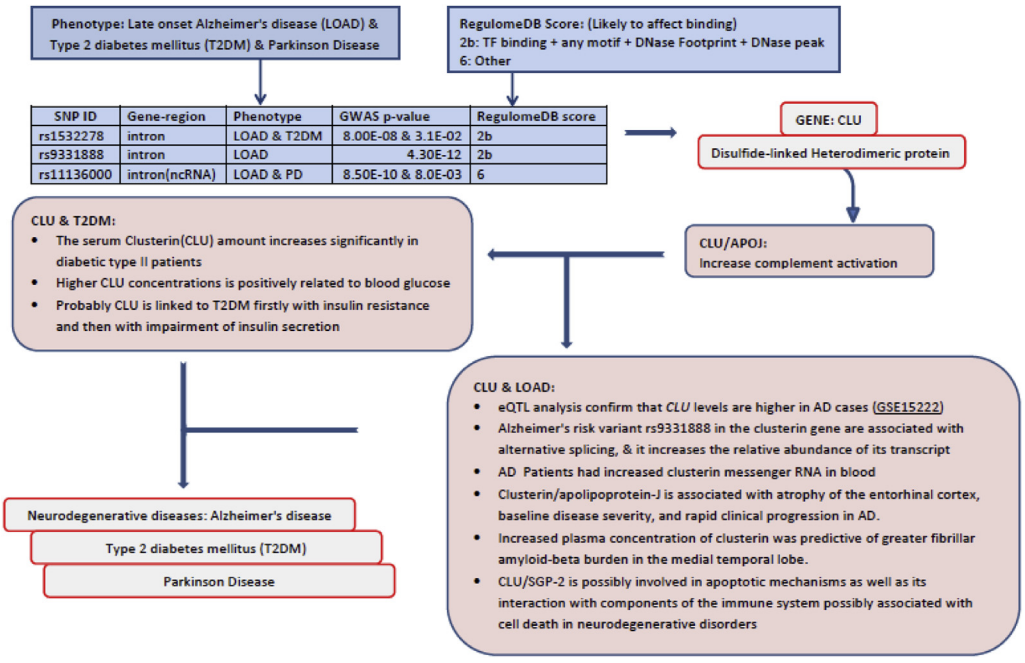


Fig. 5. Evidence-based interpretation of CLU genetic variation effect: The flowchart shows the major evidences from the biological expression language (BEL)-Model that support the mechanistic interpretation of genetic variants (single nucleotide polymorphisms) of clusterin (CLU) and links these mechanisms with disease etiology of late-onset Alzheimer's disease (LOAD) and type-2-diabetes T2DM). In diseased state, CLU is inducing the increased production of amyloid beta peptides, which is binding to insulin receptor (INSR) and inhibits the insulin-signaling pathway. Moreover, CLU is associated with an increasing risk for T2DM primarily by inducing insulin resistance and secondarily by decreasing insulin secretion. It is also increasing neuron apoptosis in diseased state.

entities). Our approach overcomes the general problem of missing values, low reproducibility, and static representation with microarray gene expression data so that differentially expressed genes detected for the same disease are often highly inconsistent and may even fail to include genes representing key causal mechanisms [45]. For instance, the functional role of neurotrophin signaling pathway in pathology of AD could be completely ignored

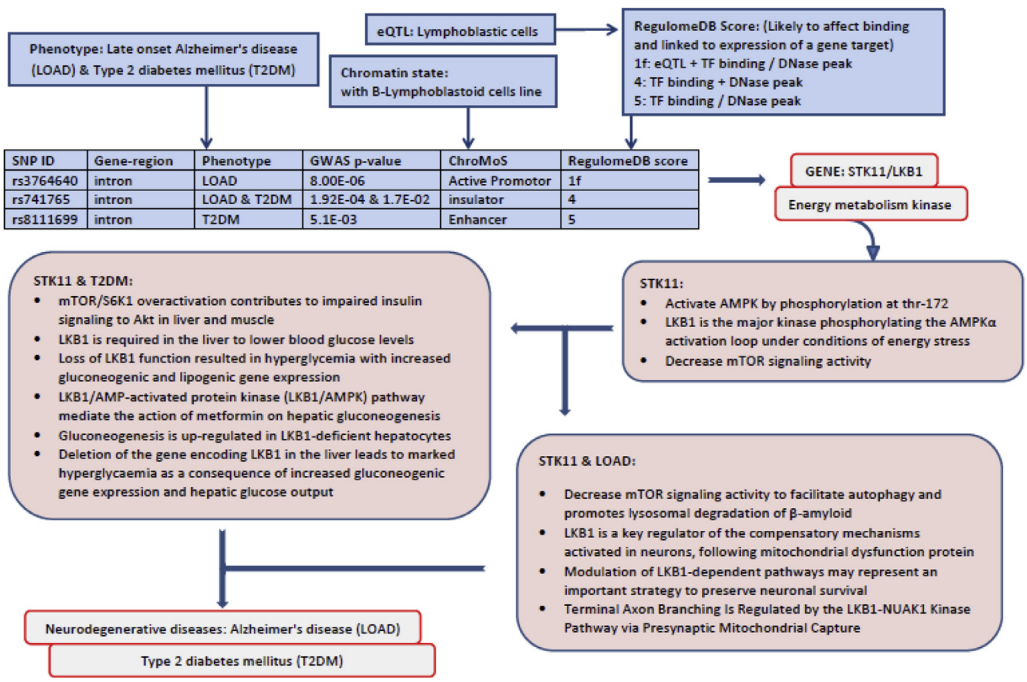


Fig. 6. Evidence-based interpretation of STK11 genetic variation effect: Cartoon-like representation flowchart of evidence encoded in the biological expression language (BEL) models that associate genetic variants (SNPs) of STK11/LKB1 gene to the putative disease etiology of late-onset Alzheimer's disease (AD) and type-2-diabetes (LOAD and T2DM).

if only gene expression values for BDNF, NGF, NTRK2, and NGRF were considered. In contrast, the power to detect this mechanism was remarkably increased when knowledge-based BEL models were used as the integrative platform for expression data.

We are of course aware of the fact that differences between the two models we generated could either reflect true pathomechanisms that differentiate the healthy and the diseased state, or—in the most trivial case—could reflect differences in the research published so far on a certain biology. We therefore emphasize that the differential model analysis is a way to generate hypotheses on possible pathomechanisms, but does not provide any proof for their true existence. Additional, independent evidence (e.g. SNP data that support the notion of an important, putative disease mechanism) and classical model validation strategies using independent data sets (e.g. RNAseq data) will help us to rapidly identify those hypotheses that merit an in-depth analysis, including experimental validation in appropriate experimental systems.

Our differential analysis of normal and disease states in AD and the additional supporting information provided evidence for the key role of amyloid-beta in switching the neurotrophin signaling pathway between cell survival and cell death. Retrospectively, we found an elegant study by Matrone et al. (2009), which lends empirical support to the role of amyloid-beta in switching from prosurvival to proapoptotic activity of the neurotrophin pathway [46]. Consistent with these results and as preclinical support for the previously mentioned hypothesis, the administration of small molecule BDNF mimetics or injection of NGF to mice models of AD has been clearly shown to result in rescue from cell death and the promotion of neuronal survival [47].

Enriched context of the APP BEL-based models with SNP data leveraged the interpretation power and allowed for linking causal effects of genetic variants to downstream molecular pathways and biological phenotypes, as exemplified for insulin resistance under AD conditions. Indeed, encoding SNPs in BEL models allows for linking SNP-associated effects to a larger functional context including biological pathways. On one hand, for most of known and statistically significant SNPs in AD GWAS results including CLU, the mode of action is not well understood but the presented BEL model in this study explains how intronic variants of CLU may increase the risk of AD through insulin resistance and increasing prevalence of T2DM. On the other hand, rare regulatory variants such as on STK11, which reside on noncoding regions of genes and are difficult to detect, have been shown to be causal for several monogenic diseases (e.g. beta-thalassemia) or modifier (e.g. sickle cell anemia) [48] but their mechanism of action is unclear. Our mechanistic models provide chains of argumentation for the causal effects of such rare regulatory variants mediated by increased levels of amyloid-beta.

It is noteworthy that BEL models go far beyond mere representation of genetic information by including downstream molecular entities and biological processes and pathways. However, BEL lacks a temporal dimension; the language has not been designed to capture kinetic information. Our future strategy to deal with the temporal dimension of Alzheimer is, to generate models representing the staging of Alzheimer by capturing the knowledge available for different stages. A first step toward stage-specific identification of biomarker candidates has already been made [49].

We envision that the application of the (qualitative) knowledge-based model provided in this study to mechanism-identification can support target identification in drug discovery and can be further enhanced by the inclusion of quantitative data. This potential has been already shown using gene expression alterations between responders and nonresponders to infliximab therapy in ulcerative colitis patients where a quantitative causal (BEL) network analysis led to the identification of a set of stratifying genes which were confirmed by their correlation with the Mayo score, a score used to diagnose patients with active ulcerative colitis [50].

5. Conclusion

Although there are clear benefits of this BEL-based, model-driven approach to understanding the complex mechanisms contributing to disease, there are some considerations for future enhancements to our models. First, given the pace of scientific research, the models need to be improved by regular update as more data and knowledge becomes available. Second, the current version of BEL describes biological interactions qualitatively and in cases where the same processes happen in both disease and normal tissue, quantitative information—when available in the literature—could allow a finer grained comparison of the diseased and normal state. Last, many studies on pathophysiology of neurodegenerative diseases like AD have been carried out in animal models, but it is not clear how well these findings are in agreement with humans. The computerization of such biological processes for representation, analysis, and comparison of interspecies mechanistic details will be a significant step forward in translational research.

Acknowledgments

The authors wish to thank Dr. Natalie. L. Catlett, Natalia Boukharov, Dr. William Hayes, and Dr. David deGraaf from Selventa Inc. for fruitful discussions and active support of this work. Furthermore, the authors acknowledge the financial support from the B-IT foundation that sponsors part of the academic work in our department. Finally, the authors would like to acknowledge the strong motivation that came from our involvement in the Neuroallianz project, a

project funded by the German Ministry of Research and Science (BMBF). Part of the research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115568 (project AETIONOMY), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2017-2013) and European Federation of Pharmaceutical Industries and Associations companies' in kind contribution.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2015.02.006.

RESEARCH IN CONTEXT

1. Systematic review: We have undertaken a comprehensive, systematic approach to capture a wide spectrum of knowledge about neuron molecular physiology in the normal state and in the diseased (Alzheimer's disease or AD) state. The knowledge gathered was used to generate two models, a "normal neuron" model and a "diseased neuron" model formalizing and representing major mechanisms underlying neuron physiology and its deregulation in disease.
2. Interpretation: With the knowledge-based modeling approach outlined in this article we substantially add to a computable, comprehensive knowledgebase in AD research. The formalism applied for the modeling supports not only sharing of knowledge, but also the identification of new candidate mechanisms underlying AD.
3. Future directions: The models we publish here are meant to provide the starting material for an ever-growing knowledgebase on AD that can be reused, expanded, and improved by the AD research community. Future directions will see an extension of the models toward epidemiological and clinical evidences, and modeling of epigenetics factors.

References

[1] Selkoe DJ. Resolving controversies on the path to Alzheimer's therapeutics. Nat Med 2011;17:1060–5.
[2] Maji SK, Perrin MH, Sawaya MR, Jessberger S, Vadodaria K, Rissman RA, et al. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. Science 2009;325:328–32.
[3] Ogishima S, Mizuno S, Kikuchi M, Miyashita A, Kuwano R, Tanaka H, et al. A map of Alzheimer's disease-signaling pathways: a hope for drug target discovery. Clin Pharmacol Ther 2013; 93:399–401.
[4] Szymanski M, Wang R, Bassett SS, Avramopoulos D. Alzheimer's risk variants in the clusterin gene are associated with alternative splicing. Transl Psychiatry 2011;1:e18.
[5] Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 2007;39(1):17–23.
[6] Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, et al. A travel guide to Cytoscape plugins. Nat Methods 2012;9:1069–76.
[7] Um SH, Frigerio F, Watanabe M. Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity. Nature 2004;431:200–5.
[8] Catlett NL, Bargnesi AJ, Ungerer S, Seagaran T, Ladd W, Elliston KO, et al. Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data. BMC Bioinformatics 2013;14:340.
[9] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
[10] Ye X, Tai W, Zhang D. The early events of Alzheimer's disease pathology: from mitochondrial dysfunction to BDNF axonal transport deficits. Neurobiol Aging 2012;33:1122.e1–112210.
[11] Poon WW, Carlos AJ, Aguilar BL, Berchtold NC, Kawano CK, Zograbyan V, et al. β-Amyloid (Aβ) oligomers impair brain-derived neurotrophic factor retrograde trafficking by down-regulating ubiquitin C-terminal hydrolase, UCH-L1. J Biol Chem 2013;288:16937–48.
[12] Geetha T, Zheng C, McGregor WC, Douglas White B, Diaz-Meco MT, Moscat J, et al. TRAF6 and p62 inhibit amyloid β-induced neuronal death through p75 neurotrophin receptor. Neurochem Int 2012; 61:1289–93.
[13] Akar CA, Wallace WC. Amyloid precursor protein modulates the interaction of nerve growth factor with p75 receptor and potentiates its activation of trkA phosphorylation. Brain Res Mol Brain Res 1998;56:125–32.
[14] Arevalo MA, Roldan PM, Chacón PJ, Rodríguez-Tebar A. Amyloid beta serves as an NGF-like neurotrophic factor or acts as a NGF antagonist depending on its concentration. J Neurochem 2009; 111:1425–33.
[15] Gilchrest Barbara A and Yaar Mina US patent: US6242416 B1 (2001).
[16] de Cubas AA, Leandro-García LJ, Schiavi F, Mancikova V, Comino-Méndez I, Inglada-Pérez L, et al. Integrative analysis of miRNA and mRNA expression profiles in pheochromocytoma and paraganglioma identifies genotype-specific markers and potentially regulated pathways. Endocr Relat Cancer 2013;20:477–93.
[17] Peng Z, Li J, Li Y, Yang X, Feng S, Han S, et al. Downregulation of miR-181b in mouse brain following ischemic stroke induces neuroprotection against ischemic injury through targeting heat shock protein A5 and ubiquitin carboxyl-terminal hydrolase isozyme L1. J Neurosci Res 2013;91:1349–62.
[18] Younesi E, Toldo L, Müller B, Friedrich CM, Novac N, Scheer A, et al. Mining biomarker information in biomedical literature. BMC Med Inform Decis Mak 2012;12:148.
[19] Akomolafe A, Beiser A, Meigs JB, Au R, Green RC, Farrer LA, et al. Diabetes mellitus and risk of developing Alzheimer disease: results from the Framingham Study. Arch Neurol 2006;63:1551–5.
[20] Luchsinger JA, Tang MX, Yaakov S, Shea S, Mayeux R. Diabetes mellitus and risk of Alzheimer disease and dementia with stroke in a multiethnic cohort. Am J Epidemiol 2001;154:635–41.
[21] Harold D, Abraham R, Hollingworth P. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009;41:1088–93.
[22] Naj AC, Jun G, Beecham GW. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 2011;43:436–41.

[23] Sagare AP, Bell RD, Zlokovic BV. Neurovascular defects and faulty amyloid-β vascular clearance in Alzheimer's disease. J Alzheimers Dis 2013;33(Suppl 1):S87–100.
[24] Nuutinen T, Suuronen T, Kauppinen A. Clusterin: a forgotten player in Alzheimer's disease. Brain Res Rev 2009;61:89–104.
[25] Shepherd CE, Mak L, Gregory GM (2012) Increases in intracellular clusterin relate to TAU pathology in Alzheimer's disease. Australian Neuroscience Society Annual Meeting, POS-WED-085.
[26] Begcevic I, Kosanam H, Martínez-Morillo E, Dimitromanolakis A, Diamandis P, Kuzmanov U, et al. Semiquantitative proteomic analysis of human hippocampal tissues from Alzheimer's disease and age-matched control brains. Clin Proteomics 2013;10:5.
[27] Michel D, Chabot JG, Moyse E, Danik M, Quirion R. Possible functions of a new genetic marker in central nervous system. The sulfated glycoprotein-2 (SGP-2). Synapse 1992;11:105–11.
[28] Karch CM, Jeng AT, Nowotny P. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. PLoS One 2012;7:e50976.
[29] Kujiraoka T, Hattori H, Miwa Y, Ishihara M, Ueno T, Ishii J, et al. Serum apolipoprotein j in health, coronary heart disease and type 2 diabetes mellitus. J Atheroscler Thromb 2006;13:314–22.
[30] Ward CW, Lawrence MC. Ligand-induced activation of the insulin receptor: a multi-step process involving structural changes in both the ligand and the receptor. Bioessays 2009;31:422–34.
[31] Jones JI, Clemmons DR. Insulin-like growth factors and their binding proteins: biological actions. Endocr Rev 1995;16:3–34.
[32] Siddle K. Signaling by insulin and IGF receptors: supporting acts and new players. J Mol Endocrinol 2011;47:R1–10.
[33] Xie L, Helmerhorst E, Taddei K, Plewright B, Van Bronswijk W, Martins R. Alzheimer's beta-amyloid peptides compete for insulin binding to the insulin receptor. J Neurosci 2002;22:RC221.
[34] Daimon M, Oizumi T, Karasawa S, Kaino W, Takase K, Tada K, et al. Association of the clusterin gene polymorphisms with type 2 diabetes mellitus. Metabolism 2011;60:815–22.
[35] López-Bermejo A, Díaz M, Morán E, de Zegher F, Ibáñez L. A single nucleotide polymorphism in STK11 influences insulin sensitivity and metformin efficacy in hyperinsulinemic girls with androgen excess. Diabetes Care 2010;33:1544–8.
[36] Barenboim M, Manke T. Genome analysis ChroMoS: an integrated web-tool for SNP classification, prioritization and functional interpretation. Bioinformatics 2013;29:2197–8.
[37] Shackelford DB, Abt E, Gerken L, Vasquez DS, Seki A, Leblanc M, et al. LKB1 inactivation dictates therapeutic response of non-small cell lung cancer to the metabolism drug phenformin. Cancer Cell 2013;23:143–58.
[38] Shaw RJ. LKB1 and AMP-activated protein kinase control of mTOR signaling and growth. Acta Physiol. (Oxf.) 2009;196:65–80.
[39] Hollingworth P, Sweet R, Sims R, Harold D, Russo G, Abraham R, et al. Genome-wide association study of Alzheimer's disease with psychotic symptoms. Mol Psychiatry 2012; 17:1316–27.
[40] Keshavarz P, Inoue H, Nakamura N, Yoshikawa T, Tanahashi T, Itakura M. Single nucleotide polymorphisms in genes encoding LKB1 (STK11), TORC2 (CRTC2) and AMPK alpha2-subunit (PRKAA2) and risk of type 2 diabetes. Mol Genet Metab 2008; 93:200–9.
[41] Courchet J, Lewis TL Jr, Lee S, Courchet V, Liou DY, Aizawa S, et al. Terminal axon branching is regulated by the LKB1-NUAK1 kinase pathway via presynaptic mitochondrial capture. Cell 2013; 153:1510–25.
[42] Carling D. LKB1: a sweet side to Peutz-Jeghers syndrome? Trends Mol Med 2006;12:144–7.
[43] Zhang B, Gaiteri C, Bodea LG. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 2013;153:707–20.
[44] Salminen A, Kaarniranta K, Haapasalo A. AMP-activated protein kinase: a potential player in Alzheimer's disease. J Neurochem 2011; 118:460–74.
[45] Zhang M, Yao C, Guo Z, Zou J, Zhang L, Xiao H, et al. Apparently low reproducibility of true differential expression discoveries in microarray studies. Bioinformatics 2008;24:2057–63.
[46] Matrone C, Marolda R, Ciafrè S, Ciotti MT, Mercanti D, Calissano P. Tyrosine kinase nerve growth factor receptor switches from prosurvival to proapoptotic activity via Abeta-mediated phosphorylation. Proc Natl Acad Sci U S A 2009;106:11358–63.
[47] Weissmiller AM, Wu C. Current advances in using neurotrophic factors to treat neurodegenerative disorders. Transl Neurodegener 2012;1:14.
[48] Li X, Montgomery SB. Detection and impact of rare regulatory variants in human disease. Front Genet 2013;4–67.
[49] Malhotra A, Younesi E, Bagewadi S, Hofmann-Apitius M. Linking hypothetical knowledge patterns to disease molecular signatures for biomarker discovery in Alzheimer's disease. Genome Med 2014;6:97.
[50] Toedter G, Li K, Sague S, Ma K, Marano C, Macoritto M, et al. Genes associated with intestinal permeability in ulcerative colitis: changes in expression following infliximab therapy. Inflamm Bowel Dis 2012; 18:1399–410.

Hindawi Publishing Corporation
Journal of Immunology Research
Volume 2015, Article ID 737168, 10 pages
http://dx.doi.org/10.1155/2015/737168

Review Article

Computational Modelling Approaches on Epigenetic Factors in Neurodegenerative and Autoimmune Diseases and Their Mechanistic Analysis

Afroza Khanam Irin,[1,2] Alpha Tom Kodamullil,[1,2] Michaela Gündel,[1,2] and Martin Hofmann-Apitius[1,2]

[1]*Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing, 53754 Sankt Augustin, Germany*
[2]*Bonn-Aachen International Center for IT, Rheinische Friedrich-Wilhelms-Universität Bonn, Dahlmannstrasse 2, 53113 Bonn, Germany*

Correspondence should be addressed to Martin Hofmann-Apitius; martin.hofmann-apitius@scai.fraunhofer.de

Received 31 July 2015; Revised 19 October 2015; Accepted 20 October 2015

Academic Editor: Francesco Pappalardo

Neurodegenerative as well as autoimmune diseases have unclear aetiologies, but an increasing number of evidences report for a combination of genetic and epigenetic alterations that predispose for the development of disease. This review examines the major milestones in epigenetics research in the context of diseases and various computational approaches developed in the last decades to unravel new epigenetic modifications. However, there are limited studies that systematically link genetic and epigenetic alterations of DNA to the aetiology of diseases. In this work, we demonstrate how disease-related epigenetic knowledge can be systematically captured and integrated with heterogeneous information into a functional context using Biological Expression Language (BEL). This novel methodology, based on BEL, enables us to integrate epigenetic modifications such as DNA methylation or acetylation of histones into a specific disease network. As an example, we depict the integration of epigenetic and genetic factors in a functional context specific to Parkinson's disease (PD) and Multiple Sclerosis (MS).

1. Introduction

In the 19th century, Gregor Mendel defined the mechanism of inheritance patterns, which laid the ground for genetics in modern biology. However, Mendel's theories could explain neither how different individuals in a population are genetically similar but exhibit different phenotypes, nor how identical twins are prone to different diseases. Recent studies confirmed that copy number variations, single nucleotide polymorphism, or any heritable changes in the DNA sequence could be a plausible additional explanation for Mendel's observation. In 1942, Waddington used the term *epigenotype* as a name for the study of causal mechanisms through which genes exhibit phenotypic effects and their adaptive interaction with the environment [1]. These epigenetic causal mechanisms involve histone modifications, DNA methylation, and abnormal RNA regulation, which can alter normal biological processes by heritable silencing of genes, although they do not cause any nucleotide sequence changes in chromosomal components [2]. Gill published the first paper describing epigenetic mechanism in drosophila egg promorphology [3]. In 1971, Tsanev and Sendov proposed the role of epigenetics in neoplastic transformation and the process of carcinogenesis [4]. Holliday reviewed the methylation of cytosine in DNA and how they are consistent to the levels of gene expression in higher organisms like human, mouse, and hamster [5]. He also illustrated that epigenetic effects are closely linked to aging such that decrease in methylation correlates with lifespan. It has later been demonstrated that epigenetic modifications are tissue-specific phenomena that can have dramatic effects on the silencing, the increase, or the reduction of the expression of genes in a given tissue. Song et al. observed variations of the methylation status in different developmental stages [6]. Additionally, Chen and

Zhang showed the risk of neonatal mortality due to maternal vascular underperfusion, which is a result of epigenetic modifications in several genes during pregnancy [7].

Several studies illustrate how nutrition and environmental factors influence epigenetic modifications. A study based on an African-American cohort demonstrated that epigenetic factors like psychological stress and social context are related to inflammation in coronary heart disease and stroke [8]. In the progression of type-2 diabetes mellitus (T2DM), Prattichizzo et al. [9] reviewed interactions between epigenetic (DNA methylation, posttranslational histone modifications, and miRNA regulation) and environmental factors (lifestyle and mainly dietary habits). Duru et al. proposed several dietary chemoprevention agents—such as Retinoids/Vitamin A, Resveratrol, EGCG/Green Tea, and Vitamin D—which act on miRNA-signalling pathways to be novel therapeutics in breast cancer [10].

It is noteworthy that environmental exposures during early stage of life can also induce persistent alterations in the epigenome, which may lead to an increased risk of disease later in life. Reviews by Van Dijk et al. and Cordero et al. investigated different epigenomics patterns in obesity during early and later stage of life [11, 12]. They elucidated the role of dietary supplements and environmental conditions on epigenetic mechanisms during the pregnancy period, which lead to the risk of obesity in offspring.

2. Epigenetics in Neurodegenerative and Autoimmune Disease

With the rising momentum of biomedical science, several studies on neurodegenerative diseases (NDDs) not only showed environmental influences on molecular and cellular changes [13, 14] but also established possible relationships between genes and the environment [15]. The major mechanisms for epigenetic alterations found in these diseases include DNA methylation, histone tail modifications, chromatin remodelling, and mechanisms regulated by small RNA molecules [16–18]. Epigenetics in neurodegenerative and autoimmune diseases are of current interest to many researchers and more recently several studies have shed light on the role of epigenetic alterations in autoimmune diseases and NDDs.

Ravaglia et al. discussed the association of folate and Vitamin B12 levels in nutritional diet with the prevalence of NDD [19]. An experiment performed on aged monkeys showed epigenetic changes in APP expression and amyloid beta level due to lead (Pb) exposure [20]. Another study by Baccarelli and Bollati explained how air-pollutants (black carbon, benzene) and toxic chemicals (arsenic, nickel, and diethylstilbestrol) alter gene expression accompanied by epigenetics changes [21]. This paper reviewed all possible metals and chemicals; those are responsible for up- or downregulation of disease specific gene such as BDNF.

Since NDDs are prevalent in the aged population, experiments conducted on NDD patients have revealed how environmental factors such as age, lifestyle, diet, and level of education influence the development of diseases and also highlighted the crosstalk of environmental factors with genes [22]. HDAC gene expression has been shown to be downregulated by Kaliman et al. due to moderate physical activities, which in turn reduce the expression of proinflammatory genes in NDDs [23]. Other than physical exercise, Nicolia et al. reviewed the role of environmental factors such as stressors (physical and behavioral), pesticides, and mental exercise causing DNA methylation in age-related diseases, specifically in AD [24]. The authors suggested that longer lifespan increases the risk of environment-induced epigenetic changes. In a detailed study [25] of epigenetics in AD, decreased DNA methylation was observed in the temporal neocortex of monozygotic AD twins. Manipulation of histone tail acetylation with HDAC inhibitors also has been investigated in several animal models of AD [26]. Martí et al. have explained a set of deregulated miRNAs that participate in altered gene expression in neurodegeneration, especially in Huntington's disease [27].

A hypothesis, namely, "hapten hypothesis," was introduced by Mintzer et al. in 2009, which describes that drugs like Penicillin and Clozapine play the role as haptens to produce antibodies against neutrophils in case of autoimmune diseases, such as Systemic Lupus Erythematosus (SLE) [28]. Uhlig et al. mentioned smoking as risk factor in addition to age and gender in another systemic autoimmune disorder, that is, Rheumatoid Arthritis (RA) [29]. Similarly, ultraviolet radiation also alters the immune mechanisms that may result in Lupus Erythematosus (LE) [30]. From the above discussion it is evident that epigenetic factors play a significant role in the context of NDD and autoimmune disease.

Although there is growing interest in epigenetics of NDDs and autoimmune diseases, only a few studies have been performed specifically on PD and MS. In fact, only a very limited number of studies deal with the functional consequences of epigenetic modifications and perturbed mechanisms leading to a particular phenotype. A systematic comparison of the number of epigenetic studies in AD, PD, and MS in the last years is shown in Figure 1(a). The graph shows that the number of scientific publications on epigenetics in PD and MS is significantly lower than the number of papers on epigenetics in AD. Figure 1(b) represents the overall trend in epigenetic studies; it becomes obvious that AD, PD, and MS represent only a minority fraction of the literature on epigenetics mechanisms, in particular when compared with the predominant indication areas arthritis, cancer, and diabetes.

3. Computational Modelling of Epigenetic Factors in a Functional Context

To represent, manipulate, and visualize large amounts of biological data from different sources, computational modelling has become an intuitive approach. Artyomov et al. proposed an "epigenetic and genetic regulatory network" that describes how transcription factors affect cellular differentiation by reprogramming embryonic cells [31]. Irrespective of any specific disease context, a computational micromodel for epigenetic mechanisms was developed by Raghavan et al.,

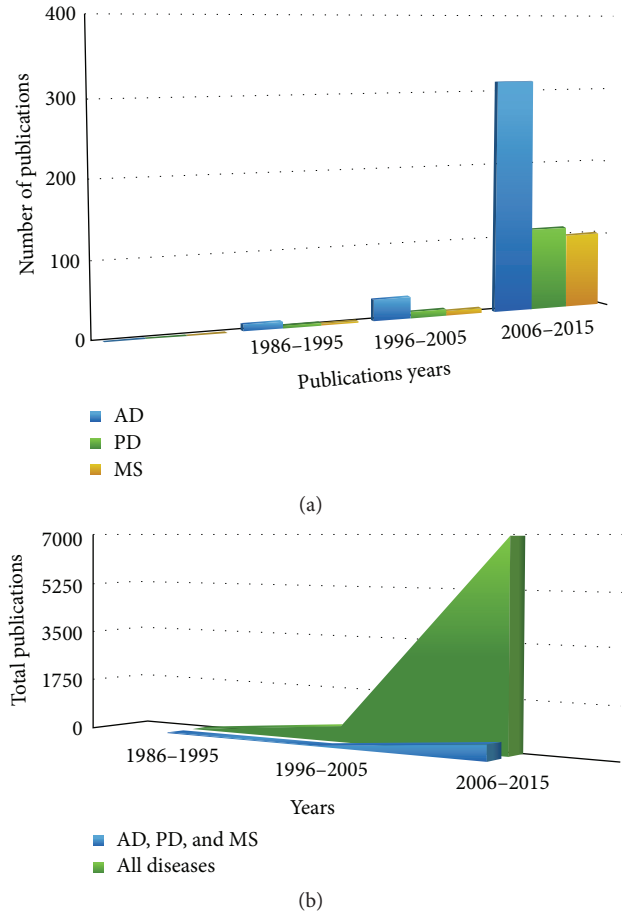


Figure 1: (a) Statistics over scientific publications around epigenetics related neurodegenerative (AD and PD), autoimmune diseases and other diseases using PubMed with queries (("Parkinson's disease") AND epigenetics), (("Alzheimer's disease") AND epigenetics), and (("Multiple Sclerosis") AND epigenetics), last accessed on 7/20/2015. In (a), blue, green, and orange coloured bars represent the total number of publications, for AD, PD, and MS, respectively. (b) This figure illustrates the trend of research on other diseases around epigenetics compared to NDD (AD and PD) and autoimmune (MS) disease, where green coloured portion representing the studies on all sorts of diseases and blue portion covers only AD, PD, and MS related researches.

demonstrating the interaction of histone modifications with DNA methylation and transcription process [32]. The model was able to identify the transcription rate when the level of DNA methylation is known.

From high throughput gene expression data of 12 human cell lines, a model integrating transcriptomic data and histone modification has been developed, called Epigenetic Regulatory Network [33], which identifies the main contributing epigenetic factors among different cell types. To facilitate the systematic integration of High Throughput Sequencing (HTS) epigenetic data, Althammer et al. have described a new computational framework. This workflow was inspired by machine learning algorithms and can be used to find alterations of epigenetic states between two given cell types [34]. Artificial Epigenetic Regulatory Network (AERN) proposed by Turner et al. has included DNA methylation and chromatin modification as the epigenetic elements in addition to genetic factors. They showed an example of how disease specific genes can be allocated in the network according to environmental changes and how gene expression regulation can be analysed within the network [35]. In a recent review paper [36], Hidden Markov Models (HMM) have been used to handle the complexity of epigenetic mechanisms, especially different patterns of DNA methylation. For autoimmune diseases, Farh et al. developed an algorithm, named "Probabilistic Identification of Causal SNPs (PICS)," which was able to find out the possibility of SNPs to be causal variants in immune cell enhancers when epigenetic modifications on that chromatin site are known [37].

Although there are algorithms that identify epigenetic modifications, there are no previous evidences describing the interpretation of functional consequences of epigenetic modifications in disease mechanisms. Here, we propose a computer-readable modelling strategy that is competent of fusing knowledge and data based information, which is capable of explaining the functional consequences of epigenetic modification in a mechanistic fashion. In this paper, we introduce the Biological Expression Language (BEL; http://www.openbel.org/) that is the main base of building models for epigenetics analysis of PD and MS.

BEL integrates literature-derived "cause and effect" relationships into network models, which can be subjected to causal analysis and used for mechanism-based hypothesis generation [38]. The semantic triple-based modelling language used here enables the application of Reverse Causal Reasoning (RCR) algorithms, which support the identification of mechanistic hypotheses from the corresponding causal network. The RCR methodology allows for investigating to what extent a knowledge-based set of triples is supported by omics data (e.g., gene expression data); the method is therefore suited for inference based on qualitatively significant data [39]. To enable a quantitative assessment and to perform comparative mechanistic analysis, another algorithm is integrated in the BEL framework: the Network Perturbation Amplitude (NPA) method. Although it uses the same network structure like RCR, its main purpose is to estimate the activity changes of a specific biological process when a pathophysiology state is compared to a nonperturbed condition [40].

Until now, BEL based network modelling approaches have been used in various applications such as early patient stratification, biomarker identification [41], and personalized drug discovery [42] in the context of cancer research by different groups. Our objective behind this computational modelling approach aims at harvesting relevant scientific knowledge from unstructured text and to systematically understand the functional impact of epigenetic modification in the context of PD and MS using BEL.

4. Role of Epigenetics in Parkinson's Disease Using BEL Models

PD is characterized by a loss of midbrain dopaminergic neurons leading to motor abnormalities and autonomic

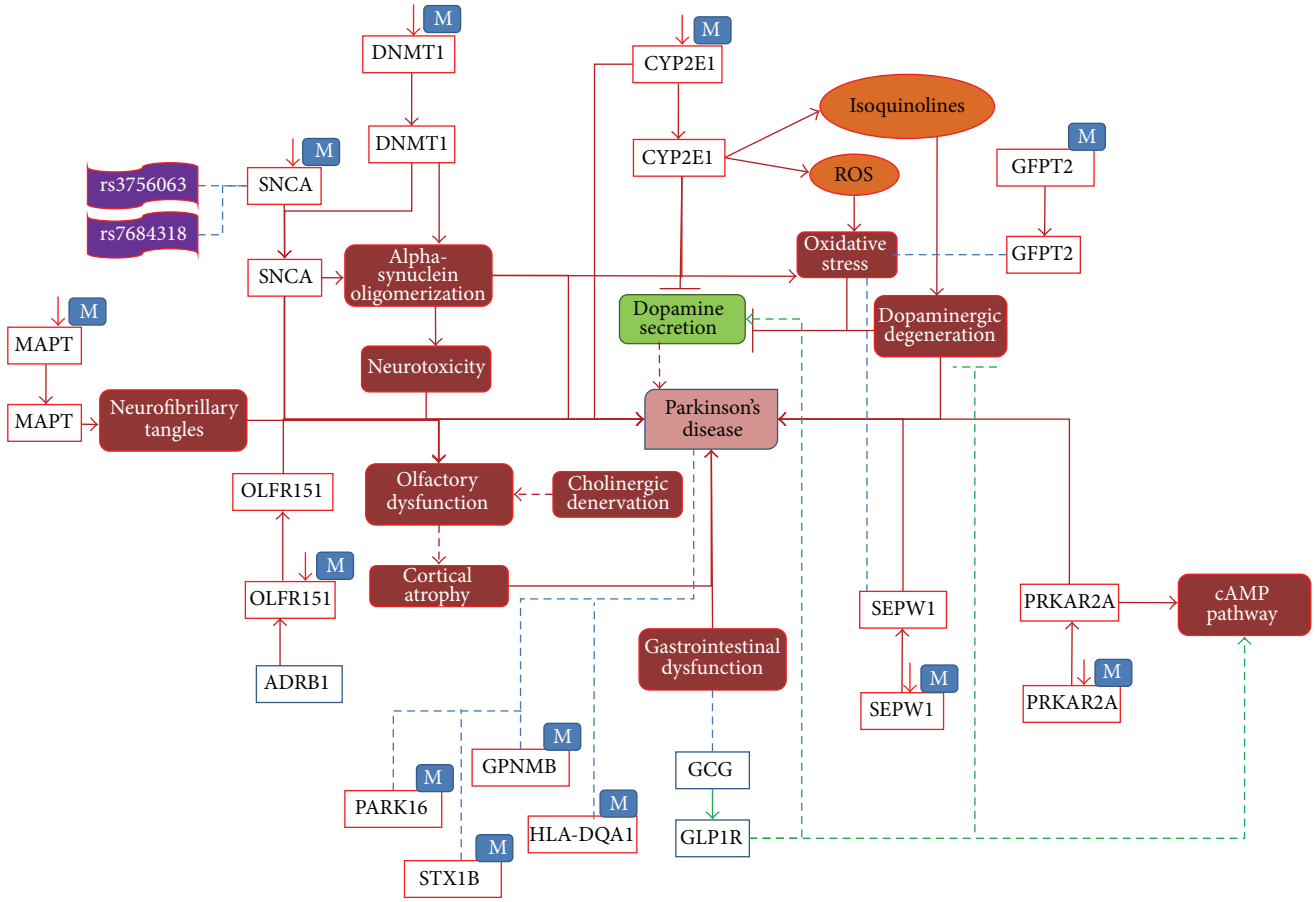


Figure 2: The role of epigenetics modification; hypomethylation around certain genes in PD. In this figure, red lines indicate the disease state interactions and green lines show normal state. Blue lines show the association between entities with unknown direction. Dotted lines are the interpretation, which needs to be further analysed. "M" associated with a gene entity denotes a methylation process and down-arrows besides represent decreased methylation.

dysfunctions [43]. Genes such as *SNCA*, *parkin*, *PINK1*, and *FBX07* have been identified to be responsible for pathophysiological mechanisms like mitochondrial damage, repair, and oxidative stress [17]. There are evidences suggesting that the above-mentioned key genes are epigenetically modified under disease conditions. For example, studies in familial as well as sporadic PD patients suggested that demethylation of the *SNCA* gene stimulates its upregulation [17, 44, 45]. Increasing amounts of *CYP2E1* have been found to promote the formation of toxic metabolites, which further degenerate the dopaminergic neurons [46]. Abnormal epigenetic modifications involved in the pathogenesis of PD have been studied by Feng et al.; in that study, detailed insights on DNA methylation and histone acetylation mechanisms and their association with the disease are reported [47].

To construct an epigenetics model for PD, we have made use of SCAIView (http://bishop.scai.fraunhofer.de/scaiview/), a literature mining environment to extract all relevant articles using the query ([*MeSH Disease*: "*Parkinson Disease*"]) *AND* ([*Parkinson Ontology*: "*Epigenetics*"]). Based on this literature mining approach, we have manually selected 78 articles, which were found to contain relevant information about PD epigenetics. The content of these publications was subsequently encoded in BEL. The model consists of 235 nodes and 407 edges representing 339 BEL statements. The nodes contain 67 proteins/genes, 21 biological processes, 6 SNPs, 3 complexes, 24 chemical entities, 26 miRNAs, and 88 other nodes representing translocation, degradation, and association functions.

As shown in Figure 2, seven representative genes, namely, *SNCA*, *MAPT*, *DNMT1*, *CYP2E1*, *OLFR151*, *PRKAR2A*, and *SEPW1*, were reported to be hypomethylated under disease conditions. In these cases, hypomethylation causes overexpression of genes that perturb normal biological processes. Increased expression of *SNCA* and *DNMT1* caused by decreased methylation of these genes results in alpha-synuclein oligomerization, which in turn causes neurotoxicity in PD [48]. Along with that, two SNPs, rs3756063 and rs7684318, were associated with hypomethylation of SNCA in PD patients. Similarly, the *CYP2E1* gene was detected to be upregulated due to (i) hypomethylation, (ii) release of isoquinolines, and (iii) Reactive Oxygen Species (ROS), which lead to dopaminergic degeneration and oxidative stress, respectively [49]. Increased neurofibrillary tangles in

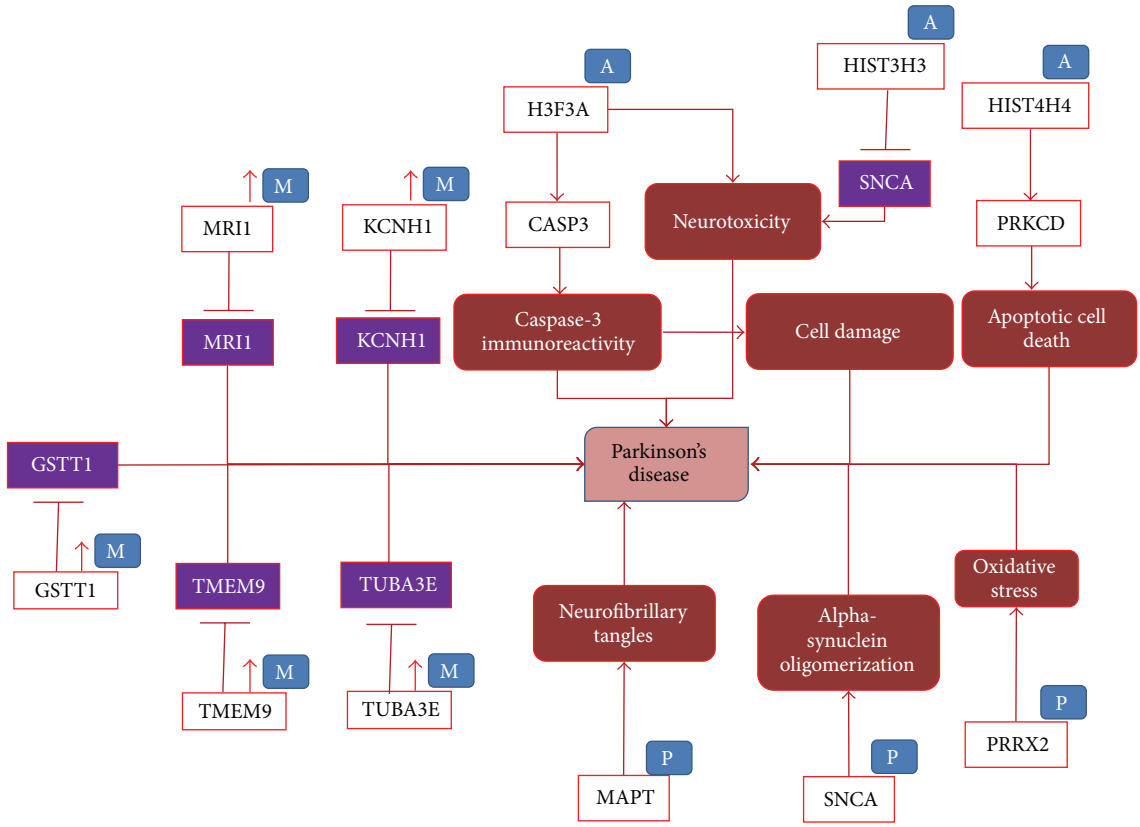


FIGURE 3: The role of epigenetics modification, hypermethylation, phosphorylation, and acetylation around certain genes in PD. In this figure also red lines indicate the disease state interactions. "M" associated with a gene entity denotes a methylation process and up-arrows besides represent an increase of methylation. "P" and "A" represent the phosphorylation and acetylation processes, respectively. Genes in purple boxes denote lower expression of genes.

PD have been reported to be linked with high expression of *MAPT* gene, as a consequence of reduced methylation [50]. Furthermore, *ADRB1* induced the hypomethylation of the *OLFR151* gene [51]. As a result, overexpression of *OLFR151* leads to olfactory dysfunction and cortical atrophy, which are early symptoms of PD [52].

GWAS and epigenomic studies suggest that *SEPW1* and *PRKAR2A* were overexpressed due to hypomethylation in PD patients [53]. However, there is lack of well-established knowledge about the functional role of *SEPW1* and *PRKAR2A* in the context of PD. We identified only one study that reports the association of *SEPW1* with PD brains [53]. Similarly, we did not find any direct biological consequences of *PRKAR2A* to play a role in the disease state. We employed a dedicated data mining approach in our model and identified the association of *PRKAR2A* with the cAMP pathway. It has been found that cAMP signal transduction pathway is stimulated by *GCG* (glucagon) [54] and its receptor *GLP1R*, which is secreted by the gastrointestinal mucosa [55]. *GLP1R* is also known to play a role in dopamine secretion and inhibiting dopaminergic degeneration [56]. Therefore we speculate that gastrointestinal dysfunction (an early symptom of PD) may result in a perturbation of the cAMP pathway and that this could be a possible mechanistic link to hypomethylation of *PRKAR2A* in PD. In addition to the above-mentioned hypomethylated genes, five more methylated genes were identified in the PD context, namely, *GFPT2*, *GPNMB*, *PARK16*, *STX1B*, and *HLA-DQA1*, where only *GFPT2* was inferred to be associated with oxidative stress [57]. These examples demonstrate that even though the analysis of high throughput data like GWAS or epigenetic studies do predict many disease-associated risk genes, no further research has been carried out to understand the functional impact of these genes.

In addition to Figure 2, we represent in our modelling approach three more highly relevant epigenetics modifications, namely, hypermethylation, phosphorylation, and acetylation (Figure 3). Five genes, *GSTT1*, *MRI1*, *KCNH1*, *TMEM9*, and *TUBA3E*, were reported to be significantly hypermethylated resulting in low expression of genes [58]. However, there were no studies describing the functional role of these genes in the PD context. In case of acetylation modification, *H3F3A*, *HIST3H3*, and *HIST4H4* were shown to be acetylated under disease conditions. Acetylated *H3F3A* increases *CASP3* activity and thereby may cause cell damage [59]. Acetylation in *HIST3H3* decreases the expression of *SNCA* leading to neurotoxicity [60], whereas *HIST4H4* acetylation induces the activity of *PRKCD*, which promotes

apoptotic cell death [59]. Phosphorylation of *MAPT*, *SNCA*, and *PRRX2* causes deposition of neurofibrillary tangles, alpha-synuclein oligomerization, and oxidative stress, respectively, in PD [50, 61].

The enlisted microRNAs in Table 1 were suggested to regulate the epigenetic modification in disease state of Parkinson. These microRNAs bind to their target and downregulate or upregulate their expression in diseased condition. For instance, *MIR34C* induces the expression of the *PARK7* gene, which in turn causes oxidative stress in PD. Some microRNAs function together (i.e., *MIR34B* and *MIR34C*) while others target individually specific genes such as *PARK7*, *PARK2*, and *TP53* to cause dysregulation in target genes, which may contribute to the disease aetiology [62].

Table 1: Role of microRNAs in PD epigenetics. 26 microRNAs have been identified that have been reported to control PD pathways. Positive and negative correlations of these microRNAs with PD mean if they are inducing or inhibiting the disease state, respectively. Also, we have enlisted the target genes for retrieved microRNAs.

Role of microRNAs in PD epigenetics		
MicroRNA	Relation to PD	Target
MIR133B	Negative correlation	*PITX3*
MIR1	Negative correlation	*TPPP, BDNF*
MIR29A	Negative correlation	—
MIR221	Negative correlation	—
MIR222	Negative correlation	—
MIR223	Negative correlation	—
MIR224	Negative correlation	—
MIR30A	Positive correlation	*SLC6A3, FGF20, GRIN1, GRIA1*
MIR16-2	Positive correlation	*FGF20*
Mir26a-2	Associated	*Gria1, Tyr*
MIR886	Positive correlation	—
MIR133B	Negative correlation	—
MIR433	Negative correlation	*FGF20*
MIR7-1	Negative correlation	—
MIR7-2	Negative correlation	—
MIR-7	Positive correlation	*SNCA*
MIR34B	Positive correlation	*PARK7, PARK2, TP53*
MIR34C	Positive correlation	*PARK7, PARK2, TP53*
MIR219A1	Negative correlation	*GRIN1, CD164*
MIR219A2	Negative correlation	*GRIN1, CD164*
MIR124-1	Positive correlation	*PPP1R13L*
Mir219a-1	Negative correlation	*Grin1*
Mir219a-2	Negative correlation	*Grin1*
Mir124a-1	Negative correlation	—
Mir124a-2	Negative correlation	—
Mir124a-3	Negative correlation	—

5. Role of Epigenetics in Multiple Sclerosis Using BEL Models

Multiple Sclerosis, a complex autoimmune disease of the central nervous system, is characterized by inflammation, demyelination, and destruction of the axons in the central nervous system [63]. Although the aetiology is not known, there is accumulating evidence that, in a cohort with genetic predisposition, environmental factors may play a key role in the development of the disease [64]. Epigenetic studies of this autoimmune disease have shown that disorders of epigenetic processes may influence chromosomal stability and gene expression, resulting in complicated syndromes [65, 66]. In a more detailed study, increased immunoreactivity for acetylated histone H3 in oligodendrocytes was found in a subset of MS samples [67]. Various microRNAs have been shown to differentially express in MS samples; particularly *MIR223* was found to be upregulated in MS patients compared to healthy controls [68]. Major epigenetic mechanisms involved in MS have been listed in a current review article [69], for example, DNA methylation, histone citrullination, and histone acetylation.

Similar to the approach taken with the PD model, we have started with a systematic literature analysis using SCAIView. We extracted information from all articles that could be retrieved with the query ([*MeSH Disease*: "*Multiple Sclerosis*"]) *AND* ([*Multiple Sclerosis Ontology*: "*Epigenetics*"]). An overall number of 75 highly relevant articles were used to build the BEL model for MS epigenetics. From this corpus of relevant literature, we have extracted 339 BEL statements to develop a network comprising 215 nodes and 536 edges. The nodes consist of 69 proteins/genes, 43 biological processes, 8 complexes, 18 chemical entities, 38 miRNAs, 8 protein families, and 31 other entities representing translocation, degradation, and association functions.

Most frequent epigenetic factors affecting MS were found to be miRNA regulation, histone citrullination, and lifestyle factors. We found 24 miRNAs that positively regulate the pathogenesis of MS and *miR23B*, *miR487B*, *miR184,* and *miR656* seem to be less expressed in the diseased context [70]. Apart from these, many epigenetics modifications like acetylation and citrullination were found in cytokines (*IFNG*, *TNF*) [71], chemokines (*CCR5*, *CCL5*, *CXCR3*, *CXCL10*, *CXCL8*, and *CXCR6*) [72], neurotrophic factors (*BDNF*, *NTF3*) [73], surface antigens (*CD8A*, *CD8B*) [74], and other genes like *GFAP*, *MBP*, *SNORD24*, and *NOTCH4*. In addition, dietary factors such as Vitamin D, intake of fruit juice, fruit/vegetables, cereal, bread, grains, and fish products reduce the risk of MS whereas intake of high energy and animal food such as fat, pork, hot dogs, and sweets increase risk of the disease (Figure 4).

6. Discussion

Epigenetics is a major mechanism that accommodates gene-expression changes in response to gene-environment interactions. In the last few decades, it has been shown that epigenetic factors play an important role in neurodegenerative as well as in autoimmune diseases. Even though there are strategies to identify new epigenetic modifications, there are very few studies, which link these alterations in DNA to the aetiology of the disease. Given the complexity and the wide variety of entities like epigenetic modifications and genetic variants, which perturb normal biological processes,

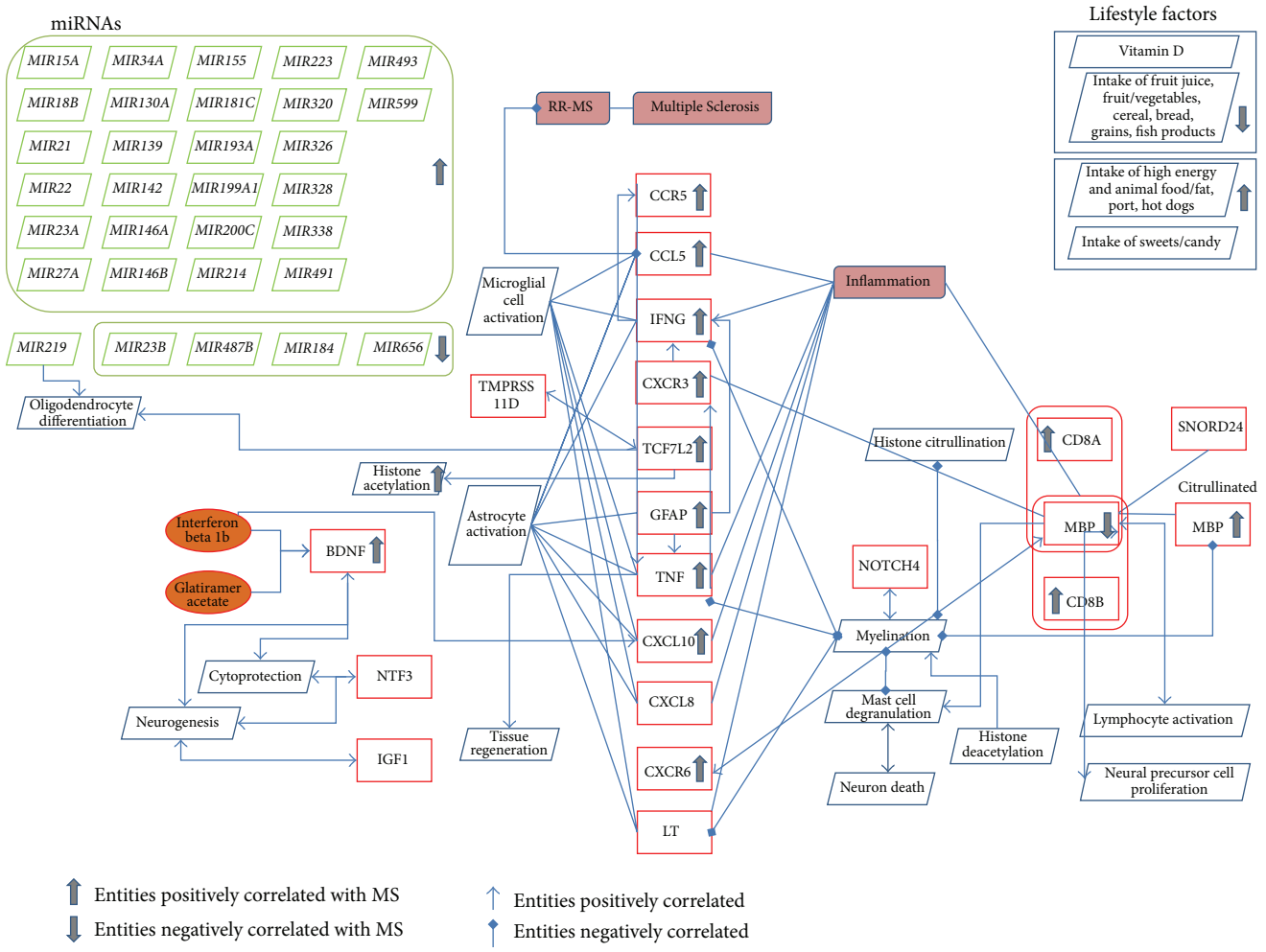


FIGURE 4: Model around epigenetic factors in Multiple Sclerosis. The figure shows the different epigenetic factors which regulate MS pathology including miRNAs represented in green colour boxes and genes in red colour boxes, chemicals in orange colour, and bioprocess in blue rhombus shape.

we need new strategies to integrate data driven and knowledge driven approaches to unravel the mechanisms behind these complex diseases. We demonstrated that it is possible to collectively capture disease-related, epigenetic knowledge and integrate it into a functional context using the modelling language BEL. An adaptation of the BEL syntax enables us to integrate epigenetic modification information like methylation (hypo and hyper), acetylation, phosphorylation, and miRNAs regulation into a specific disease network. In addition to these mechanisms, we have also included the role of many environmental factors such as food habit and obesity to the model which are responsible for the epigenetic modifications.

Although fewer studies related to PD and MS around epigenetics have been published until now, we tried to integrate all available knowledge from the scientific literature. In the case of PD, the main genes which are epigenetically regulated through methylation are *SNCA*, *PARK6*, *CYP2E1*, *PINK1*, *BDNF*, *FGF*, *MAPT*, *MTHFR*, *OLFR 151*, *PARK16*, *PARK2*, *PARK7*, *TPPP*, *PDE4D*, and *METRNL*. Also we have found acetylation in *H3F3A*, *HIST3H3*, and *HIST4H4* genes and phosphorylation in *MAPT*, *SNCA*, and *PRRX2* genes as major epigenetic modifications in PD along with miRNA regulation. Similarly for MS, we have found several citrullinated or acetylated cytokines, chemokines, transcription factors, neurotrophins, and many dietary factors, which can influence disease processes.

Some of the genes identified are well studied, but for others still an in-depth analysis is needed. Since there are no studies published on these novel candidates derived from data driven approaches, we were not able to link the functional impact of epigenetic modifications to the disease aetiology. For instance, there are about 30 GWAS studies associating the *PARK16* gene with PD, but no detailed information about the functional context of *PARK16* in the pathophysiology of PD exists in the literature. We observe a clear bias towards well-known candidate genes like *SNCA* for PD and *MBP* for MS; in order to overcome this bias, dedicated effort towards investigating the role of the new candidate genes and related bioprocesses is required.

Although BEL has the capability to integrate different biological entities and modifications at the levels of proteins, the current version of BEL is not efficient in representing epigenetic modifications at gene level, so that it is not yet possible to reason over epigenetic effects automatically (e.g., using RCR). It is obvious that we need to extend the syntax of the modelling language in order to formally represent this type of variation and develop algorithms that assess the functional impact based on biological network models.

Abbreviations

BEL: Biological Expression Language
PD: Parkinson's disease
MS: Multiple Sclerosis
AD: Alzheimer's disease
NDDs: Neurodegenerative diseases
T2DM: Type-2 diabetes mellitus
DNA: Deoxyribonucleic acid
RNA: Ribonucleic acid
HDAC: Histone deacetylase
APP: Amyloid Precursor Protein
SLE: Systemic Lupus Erythematosus
RA: Rheumatoid Arthritis
HTS: High throughput sequencing
AGRN: Artificial Gene Regulatory Network
AERN: Artificial Epigenetic Regulatory Network
HMM: Hidden Markov Model
SNP: Single nucleotide polymorphism
ROS: Reactive Oxygen Species
GWAS: Genome-Wide Association Study
cAMP: Cyclic adenosine $3'$,$5'$-monophosphate
miRNA: MicroRNA.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors wish to thank Professor Dr. Ullrich Wüllner from Universitätsklinikum Bonn for fruitful discussions and active support for this work. Furthermore, the authors acknowledge the financial support from the B-IT foundation that sponsors part of the academic work in their department. Finally, the authors would like to acknowledge the strong motivation that came from their involvement in the Neuroallianz project, a project funded by the German Ministry of Research and Science (BMBF). Part of the research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under Grant Agreement no. 115568 (project AETIONOMY), resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2017-2013) and European Federation of Pharmaceutical Industries and Associations companies in kind contribution.

References

[1] C. H. Waddington, "The epigenotype. 1942," *International Journal of Epidemiology*, vol. 41, no. 1, pp. 10–13, 2012.

[2] G. Egger, G. Liang, A. Aparicio, and P. A. Jones, "Epigenetics in human disease and prospects for epigenetic therapy," *Nature*, vol. 429, no. 6990, pp. 457–463, 2004.

[3] K. S. Gill, "Epigenetics of the promorphology of the egg in *Drosophila Melanogaster*," *Journal of Experimental Zoology*, vol. 155, no. 1, pp. 91–104, 1964.

[4] R. Tsanev and B. Sendov, "An epigenetic mechanism for carcinogenesis," *Zeitschrift für Krebsforschung und Klinische Onkologie*, vol. 76, no. 4, pp. 299–319, 1971.

[5] R. Holliday, "The inheritance of epigenetic defects," *Science*, vol. 238, no. 4824, pp. 163–170, 1987.

[6] F. Song, S. Mahmood, S. Ghosh et al., "Tissue specific differentially methylated regions (TDMR): changes in DNA methylation during development," *Genomics*, vol. 93, no. 2, pp. 130–139, 2009.

[7] M. Chen and L. Zhang, "Epigenetic mechanisms in developmental programming of adult disease," *Drug Discovery Today*, vol. 16, no. 23-24, pp. 1007–1018, 2011.

[8] K. L. Saban, H. L. Mathews, H. A. de Von, and L. W. Janusek, "Epigenetics and social context: implications for disparity in cardiovascular disease," *Aging and Disease*, vol. 5, no. 5, pp. 346–355, 2014.

[9] F. Prattichizzo, A. Giuliani, A. Ceka et al., "Epigenetic mechanisms of endothelial dysfunction in type 2 diabetes," *Clinical Epigenetics*, vol. 7, article 56, 2015.

[10] N. Duru, R. Gernapudi, G. Eades, R. Eckert, and Q. Zhou, "Epigenetic regulation of miRNAs and breast cancer stem cells," *Current Pharmacology Reports*, vol. 1, no. 3, pp. 161–169, 2015.

[11] S. J. Van Dijk, P. L. Molloy, H. Varinli et al., "Epigenetics and human obesity," *International Journal of Obesity*, vol. 39, no. 1, pp. 85–97, 2015.

[12] P. Cordero, J. Li, and J. A. Oben, "Epigenetics of obesity: beyond the genome sequence," *Current Opinion in Clinical Nutrition & Metabolic Care*, vol. 18, no. 4, pp. 361–366, 2015.

[13] T. Palomo, T. Archer, R. J. Beninger, and R. M. Kostrzewa, "Gene-environment interplay in neurogenesis and neurodegeneration," *Neurotoxicity Research*, vol. 6, no. 6, pp. 415–434, 2004.

[14] T. L. Spires and A. J. Hannan, "Nature, nurture and neurology: gene-environment interactions in neurodegenerative disease. FEBS Anniversary Prize Lecture delivered on 27 June 2004 at the 29th FEBS Congress in Warsaw," *FEBS Journal*, vol. 272, no. 10, pp. 2347–2361, 2005.

[15] F. Coppedè, M. Mancuso, G. Siciliano, L. Migliore, and L. Murri, "Genes and the environment in neurodegeneration," *Bioscience Reports*, vol. 26, no. 5, pp. 341–367, 2006.

[16] S. C. Marques, C. R. Oliveira, C. M. Pereira, and T. F. Outeiro, "Epigenetics in neurodegeneration: a new layer of complexity," *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, vol. 35, no. 2, pp. 348–355, 2011.

[17] F. Coppedè, "Genetics and epigenetics of Parkinson's disease," *The Scientific World Journal*, vol. 2012, Article ID 489830, 12 pages, 2012.

[18] J. M. Greer and P. A. McCombe, "The role of epigenetic mechanisms and processes in autoimmune disorders," *Biologics*, vol. 6, pp. 307–327, 2012.

[19] G. Ravaglia, P. Forti, F. Maioli et al., "Homocysteine and folate as risk factors for dementia and Alzheimer disease," *The American Journal of Clinical Nutrition*, vol. 82, no. 3, pp. 636–643, 2005.

[20] J. Wu, M. R. Basha, B. Brock et al., "Alzheimer's disease (AD)-like pathology in aged monkeys after infantile exposure to environmental metal lead (Pb): evidence for a developmental origin and environmental link for AD," *The Journal of Neuroscience*, vol. 28, no. 1, pp. 3–9, 2008.

[21] A. Baccarelli and V. Bollati, "Epigenetics and environmental chemicals," *Current Opinion in Pediatrics*, vol. 21, no. 2, pp. 243–251, 2009.

[22] I. A. Qureshi and M. F. Mehler, "Advances in epigenetics and epigenomics for neurodegenerative diseases," *Current Neurology and Neuroscience Reports*, vol. 11, no. 5, pp. 464–473, 2011.

[23] P. Kaliman, M. J. Álvarez-López, M. Cosín-Tomás, M. A. Rosenkranz, A. Lutz, and R. J. Davidson, "Rapid changes in histone deacetylases and inflammatory gene expression in expert meditators," *Psychoneuroendocrinology*, vol. 40, no. 1, pp. 96–107, 2014.

[24] V. Nicolia, M. Lucarelli, and A. Fuso, "Environment, epigenetics and neurodegeneration: focus on nutrition in Alzheimer's disease," *Experimental Gerontology*, vol. 68, pp. 8–12, 2015.

[25] D. Mastroeni, A. Grover, E. Delvaux, C. Whiteside, P. D. Coleman, and J. Rogers, "Epigenetic changes in Alzheimer's disease: decrements in DNA methylation," *Neurobiology of Aging*, vol. 31, no. 12, pp. 2025–2037, 2010.

[26] Y. I. Francis, M. Fà, H. Ashraf et al., "Dysregulation of histone acetylation in the APP/PS1 mouse model of Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 18, no. 1, pp. 131–139, 2009.

[27] E. Martí, L. Pantano, M. Bañez-Coronel et al., "A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing," *Nucleic Acids Research*, vol. 38, no. 20, pp. 7219–7235, 2010.

[28] D. M. Mintzer, S. N. Billet, and L. Chmielewski, "Drug-induced hematologic syndromes," *Advances in Hematology*, vol. 2009, Article ID 495863, 11 pages, 2009.

[29] T. Uhlig, K. B. Hagen, and T. K. Kvien, "Current tobacco smoking, formal education, and the risk of rheumatoid arthritis," *Journal of Rheumatology*, vol. 26, no. 1, pp. 47–54, 1999.

[30] A. Kuhn and S. Beissert, "Photosensitivity in lupus erythematosus," *Autoimmunity*, vol. 38, no. 7, pp. 519–529, 2005.

[31] M. N. Artyomov, A. Meissner, and A. K. Chakraborty, "A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency," *PLoS Computational Biology*, vol. 6, no. 5, Article ID e1000785, 2010.

[32] K. Raghavan, H. J. Ruskin, D. Perrin, F. Goasmat, and J. Burns, "Computational micromodel for epigenetic mechanisms," *PLoS ONE*, vol. 5, no. 11, Article ID e14031, 2010.

[33] L. Y. Wang, P. Wang, M. J. Li et al., "EpiRegNet: constructing epigenetic regulatory network from high throughput gene expression data for humans," *Epigenetics*, vol. 6, no. 12, pp. 1505–1512, 2011.

[34] S. Althammer, A. Pagès, and E. Eyras, "Predictive models of gene regulation from high-throughput epigenomics data," *Comparative and Functional Genomics*, vol. 2012, Article ID 284786, 13 pages, 2012.

[35] A. P. Turner, M. A. Lones, L. A. Fuente, S. Stepney, L. S. D. Caves, and A. M. Tyrrell, "The incorporation of epigenetics in artificial gene regulatory networks," *BioSystems*, vol. 112, no. 2, pp. 56–62, 2013.

[36] K.-E. Lee and H.-S. Park, "A review of three different studies on hidden markov models for epigenetic problems: a computational perspective," *Genomics & Informatics*, vol. 12, no. 4, pp. 145–150, 2014.

[37] K. K.-H. Farh, A. Marson, J. Zhu et al., "Genetic and epigenetic fine mapping of causal autoimmune disease variants," *Nature*, vol. 518, no. 7539, pp. 337–343, 2015.

[38] A. T. Kodamullil, E. Younesi, M. Naz, S. Bagewadi, and M. Hofmann-Apitius, "Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis," *Alzheimer's & Dementia*, 2015.

[39] N. L. Catlett, A. J. Bargnesi, S. Ungerer et al., "Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data," *BMC Bioinformatics*, vol. 14, article 340, 2013.

[40] F. Martin, T. M. Thomson, A. Sewer et al., "Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks," *BMC Systems Biology*, vol. 6, article 54, 2012.

[41] D. Laifenfeld, D. A. Drubin, N. L. Catlett et al., "Early patient stratification and predictive biomarkers in drug discovery and development: a case study of ulcerative colitis anti-TNF therapy," *Advances in Experimental Medicine and Biology*, vol. 736, pp. 645–653, 2012.

[42] D. A. Fryburg, D. H. Song, and D. De Graaf, "Early patient stratification is critical to enable effective and personalised drug discovery and development," *Drug Discovery World*, vol. 12, no. 3, pp. 47–56, 2011.

[43] B. Thomas and M. F. Beal, "Molecular insights into Parkinson's disease," *F1000 Medicine Reports*, vol. 3, article 7, 2011.

[44] A. Jowaed, I. Schmitt, O. Kaut, and U. Wüllner, "Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains," *The Journal of Neuroscience*, vol. 30, no. 18, pp. 6355–6359, 2010.

[45] L. Matsumoto, H. Takuma, A. Tamaoka et al., "CpG demethylation enhances alpha-synuclein expression and affects the pathogenesis of Parkinson's disease," *PLoS ONE*, vol. 5, no. 11, Article ID e15522, 2010.

[46] A. G. Riedl, P. M. Watts, P. Jenner, and C. D. Marsden, "P450 enzymes and Parkinson's disease: the story so far," *Movement Disorders*, vol. 13, no. 2, pp. 212–220, 1998.

[47] Y. Feng, J. Jankovic, and Y.-C. Wu, "Epigenetic mechanisms in Parkinson's disease," *Journal of the Neurological Sciences*, vol. 349, no. 1-2, pp. 3–9, 2015.

[48] O. W. Wan and K. K. K. Chung, "The role of alpha-synuclein oligomerization and aggregation in cellular and animal models of Parkinson's disease," *PLoS ONE*, vol. 7, no. 6, Article ID e38545, 2012.

[49] O. Kaut, I. Schmitt, and U. Wüllner, "Genome-scale methylation analysis of Parkinson's disease patients' brains reveals DNA hypomethylation and increased mRNA expression of cytochrome P450 2E1," *Neurogenetics*, vol. 13, no. 1, pp. 87–91, 2012.

[50] M. J. Devine and P. A. Lewis, "Emerging pathways in genetic Parkinson's disease: tangles, Lewy bodies and LRRK2," *The FEBS Journal*, vol. 275, no. 23, pp. 5748–5757, 2008.

[51] C. Hague, M. A. Uberti, Z. Chen et al., "Olfactory receptor surface expression is driven by association with the beta2-adrenergic receptor," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 37, pp. 13672–13676, 2004.

[52] E.-Y. Lee, P. J. Eslinger, G. Du, L. Kong, M. M. Lewis, and X. Huang, "Olfactory-related cortical atrophy is associated with olfactory dysfunction in Parkinson's disease," *Movement Disorders*, vol. 29, no. 9, pp. 1205–1208, 2014.

[53] P. Desplats, B. Spencer, E. Coffee et al., "Alpha-synuclein sequesters Dnmt1 from the nucleus: a novel mechanism for epigenetic alterations in Lewy body diseases," *The Journal of Biological Chemistry*, vol. 286, no. 11, pp. 9031–9037, 2011.

[54] P. Viitala, K. Posti, A. Lindfors, O. Pelkonen, and H. Raunio, "cAMP mediated upregulation of CYP2A5 in mouse hepatocytes," *Biochemical and Biophysical Research Communications*, vol. 280, no. 3, pp. 761–767, 2001.

[55] Y. Fujii, N. Osaki, T. Hase, and A. Shimotoyodome, "Ingestion of coffee polyphenols increases postprandial release of the active glucagon-like peptide-1 (GLP-1(7–36)) amide in C57BL/6J mice," *Journal of Nutritional Science*, vol. 4, article e9, 9 pages, 2015.

[56] Y. Li, T. Perry, M. S. Kindy et al., "GLP-1 receptor stimulation preserves primary cortical and dopaminergic neurons in cellular and rodent models of stroke and Parkinsonism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 4, pp. 1285–1290, 2009.

[57] P. Prasad, A. K. Tiwari, K. M. P. Kumar et al., "Association analysis of ADPRT1, AKR1B1, RAGE, GFPT2 and PAI-1 gene polymorphisms with chronic renal insufficiency among Asian Indians with type-2 diabetes," *BMC Medical Genetics*, vol. 11, article 52, 2010.

[58] E. Masliah, W. Dumaop, D. Galasko, and P. Desplats, "Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes," *Epigenetics*, vol. 8, no. 10, pp. 1030–1038, 2013.

[59] C. Song, A. Kanthasamy, V. Anantharam, F. Sun, and A. G. Kanthasamy, "Environmental neurotoxic pesticide increases histone acetylation to promote apoptosis in dopaminergic neuronal cells: relevance to epigenetic mechanisms of neurodegeneration," *Molecular Pharmacology*, vol. 77, no. 4, pp. 621–632, 2010.

[60] I. F. Harrison and D. T. Dexter, "Epigenetic targeting of histone deacetylase: therapeutic potential in Parkinson's disease?" *Pharmacology and Therapeutics*, vol. 140, no. 1, pp. 34–52, 2013.

[61] D. Qu, J. Rashidian, M. P. Mount et al., "Role of Cdk5-Mediated Phosphorylation of Prx2 in MPTP Toxicity and Parkinson's Disease," *Neuron*, vol. 55, no. 1, pp. 37–52, 2007.

[62] E. Miñones-Moyano, S. Porta, G. Escaramís et al., "MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function," *Human Molecular Genetics*, vol. 20, no. 15, pp. 3067–3078, 2011.

[63] D. A. Umphred, *Neurological Rehabilitation*, Mosby, St. Louis, Mo, USA, 2001.

[64] M. Iridoy Zulet, L. Pulido Fontes, T. Ayuso Blanco, F. Lacruz Bescos, and M. Mendioroz Iriarte, "Epigenetic changes in neurology: DNA methylation in multiple sclerosis," *Neurología*, 2015.

[65] S. Ruhrmann, P. Stridh, L. Kular, and M. Jagodic, "Genomic imprinting: a missing piece of the multiple sclerosis puzzle?" *The International Journal of Biochemistry & Cell Biology*, vol. 67, pp. 49–57, 2015.

[66] Z. Zhang and R. Zhang, "Epigenetics in autoimmune diseases: pathogenesis and prospects for therapy," *Autoimmunity Reviews*, vol. 14, no. 10, pp. 854–863, 2015.

[67] X. Pedre, F. Mastronardi, W. Bruck, G. López-Rodas, T. Kuhlmann, and P. Casaccia, "Changed histone acetylation patterns in normal-appearing white matter and early multiple sclerosis lesions," *The Journal of Neuroscience*, vol. 31, no. 9, pp. 3435–3445, 2011.

[68] A. Keller, P. Leidinger, J. Lange et al., "Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls," *PLoS ONE*, vol. 4, no. 10, Article ID e7440, 2009.

[69] C. İ. Küçükali, M. Kürtüncü, A. Çoban, M. Çebi, and E. Tüzün, "Epigenetics of multiple sclerosis: an updated review," *NeuroMolecular Medicine*, vol. 17, no. 2, pp. 83–96, 2015.

[70] A. Junker, M. Krumbholz, S. Eisele et al., "MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47," *Brain*, vol. 132, no. 12, pp. 3342–3352, 2009.

[71] M. Sospedra and R. Martin, "Immunology of multiple sclerosis," *Annual Review of Immunology*, vol. 23, pp. 683–747, 2005.

[72] T. L. Sørensen, M. Tani, J. Jensen et al., "Expression of specific chemokines and chemokine receptors in the central nervous system of multiple sclerosis patients," *Journal of Clinical Investigation*, vol. 103, no. 6, pp. 807–815, 1999.

[73] R. Gandhi, A. Laroni, and H. L. Weiner, "Role of the innate immune system in the pathogenesis of multiple sclerosis," *Journal of Neuroimmunology*, vol. 221, no. 1-2, pp. 7–14, 2010.

[74] Y. C. Q. Zang, S. Li, V. M. Rivera et al., "Increased $CD8^+$ cytotoxic T cell responses to myelin basic protein in multiple sclerosis," *Journal of Immunology*, vol. 172, no. 8, pp. 5120–5127, 2004.

Summary

In the AD research field, the number of research articles is extensive, and it is unfeasible for a researcher to read and retain all the relevant knowledge from the vast number of articles produced. Therefore, it is essential to have an integrative model which can store knowledge in a single format and retrieve relevant information at any time. In the preceding chapter, I showed how BEL is used to build cause and effect based computable models of AD and of normal physiological functions in the human brain. The AD models depicted in the previous publications are comprised of biological entities ranging from the genomic to the phenotypic level.

The primary aim of the modelling described above is to gain insight and understanding of the underlying mechanisms of AD. By comparing the normal healthy model with the AD model, I have identified an early mechanism by which the neurotrophic signaling pathway is perturbed. Based on the comparative analysis, I have also made apparent a publication bias exists in AD research towards the amyloid hypothesis in the AD context, as well as demonstrated that there are fewer studies which focus on the normal function of the brain. Additional drug-targetable mechanisms in AD (more than one hundred) were identified based on this work (see additional publications: Domingo-Fernández D, Kodamullil AT et al, (2017)).

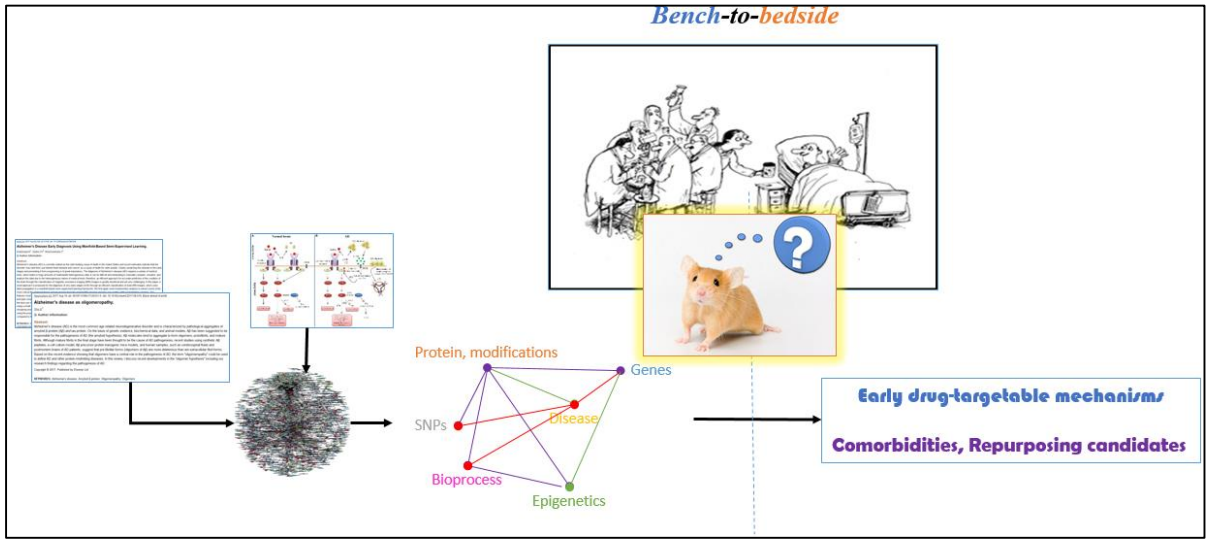
In addition, BEL models were used during the course of this work to include epigenetic modifications and to find the functional consequences of epigenetic modifications in the context of AD. These models also allowed the extraction of shared mechanisms between AD and T2DM and demonstrated the functional role of SNPs in comorbidity association between

these two diseases. This finding allowed us to explore options for the repurposing of T2DM drug candidates in AD as well as to reconsider, whether T2DM is a risk factor for the development of AD (see additional publications: Karki R, Kodamullil AT et.al (2017)).

As a follow up of my initial modeling and analysis work, Emon MA, Kodamullil AT et.al (2017), (see additional publications) annotated the model with drug-target information and proposed two repositioning candidates for Alzheimer's disease and one for amyotrophic lateral sclerosis (ALS).

Chapter 2

Analyzing the predictive power of the mouse model for drug efficacy in human



Introduction

Although there is an overwhelming amount of research conducted in the AD field, the success rate of clinical trials in this field is less than 0.5 percent (Cummings 2014). At present, AD is not curable and the approved drugs in AD are limited to symptom reduction. Animal models, especially mouse models, contribute significantly in deciphering the underlying mechanism of diseases. Nevertheless, it is debatable as to what extent genomic responses of humans can be translated to the mouse, particularly due to the discordance of clinical trials in humans and pre-clinical drug assessments (Seok 2013), (Warren 2015). In this work, we have performed a comparative, functional analysis of on neuro-inflammatory pathways in mouse and human. We also show marked divergence in drug-targeted pathways between the species, based on a discontinued AD drug but approved for different disease.

References

Cummings 2014: Cummings, Jeffrey L and Morstorf, Travis and Zhong, Kate. 2014. „Alzheimer's disease drug-development pipeline: few candidates, frequent failures.“ *Alzheimer's research \& therapy* 6: 37

Seok 2013: Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, Richards DR, McDonald-Smith GP, Gao H, Hennessy L, Finnerty CC, López CM, Honari S, Moore EE, Minei JP, Cuschieri J, Bankey PE, Johnson JL, Sperry J, Nathens AB, Billiar TR, West MA, Jeschke MG, Klein MB, Gamelli RL, Gibran NS, Brownstein BH, Miller-Graziano C, Calvano SE, Mason PH, Cobb JP, Rahme LG, Lowry SF, Maier RV, Moldawer LL, Herndon DN, Davis RW, Xiao W, Tompkins RG, Inflammation and Host Response to Injury, Large Scale Collaborative Research Program. "Genomic responses in mouse models poorly mimic human inflammatory diseases." *Proc Natl Acad Sci U S A. 110 :3507-12*

Warren 2015: Warren HS, Tompkins RG, Moldawer LL, Seok J, Xu W, Mindrinos MN, Maier RV, Xiao W, Davis RW. "Mice are not men." *Proc Natl Acad Sci U S A; 112(4):E345.*

Journal of Alzheimer's Disease 59 (2017) 1045–1055
DOI 10.3233/JAD-170255
IOS Press

Of Mice and Men: Comparative Analysis of Neuro-Inflammatory Mechanisms in Human and Mouse Using Cause-and-Effect Models

Alpha Tom Kodamullil[a,b], Anandhi Iyappan[a,b,1], Reagon Karki[a,b,1], Sumit Madan[a,b], Erfan Younesi[a] and Martin Hofmann-Apitius[a,b,*]
[a]*Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin, Germany*
[b]*Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn-Aachen International Center for IT, Bonn, Germany*

Accepted 6 June 2017

Abstract. Perturbance in inflammatory pathways have been identified as one of the major factors which leads to neurodegenerative diseases (NDD). Owing to the limited access of human brain tissues and the immense complexity of the brain, animal models, specifically mouse models, play a key role in advancing the NDD field. However, many of these mouse models fail to reproduce the clinical manifestations and end points of the disease. NDD drugs, which passed the efficacy test in mice, were repeatedly not successful in clinical trials. There are numerous studies which are supporting and opposing the applicability of mouse models in neuroinflammation and NDD. In this paper, we assessed to what extend a mouse can mimic the cellular and molecular interactions in humans at a mechanism level. Based on our mechanistic modeling approach, we investigate the failure of a neuroinflammation targeted drug in the late phases of clinical trials based on the comparative analyses between the two species.

Keywords: Alzheimer's disease, human, mice, neuroinflammation

INTRODUCTION

Neuroinflammation is the hallmark of almost all neurodegenerative diseases (NDDs) including Alzheimer's disease (AD) [1]. Aggregated amyloid-β (Aβ) peptides are believed to trigger the innate immune response through microglial and astroglial cells, which may lead to exacerbation of the disease [2]. Studies on early stage of AD as well as rodent models suggest that immune actions alone are sufficient to cause AD-like pathology and can precede tau and amyloid pathology in the brain [3]. As a consequence, neuroinflammation in AD has been proposed as an attractive target for therapeutic modulation and prevention [4]. Modulation of neuroinflammation for drug target- or biomarker identification requires extensive use of rodent models of AD to study molecular drivers of inflammation and various disease phenotypes associated to it [5]. Despite the availability of different mouse models representing APP mutations or tauopathy, the results of neuroinflammation modulation in these models have been divergent, suggesting that currently available mouse models do not accurately reflect human AD pathology [6]. For instance, conventional transgenic models of AD, which are routinely used for preclinical studies, have been shown to incompletely mirror the inflammatory response seen in AD human

[1]Authors contributed equally to this work.

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

ISSN 1387-2877/17/$35.00

brains [7]. Work of Seok et al. in 2013 that reported on poor recapitulation of genomic responses of human inflammatory diseases in mouse models [8] stimulated the old debate [9] as to whether animal models can reliably inform human diseases. However, statistical re-evaluation of Seok et al. results by Takao and Miyakawa in 2014 suggested that correlations between gene expression patterns from mouse models and human conditions were stronger than reported originally [10]. Warren et al. re-confirmed essential differences between these two species at molecular level by showing that mouse models mimicked only 12% of the genes deregulated in human conditions [11]. Beside these inter-species differences between inflammatory responses in human and mouse at molecular level, a similar significant difference also exists at the brain anatomical level, including greater size, higher lobular organization, more developed sulci and gyri, and larger amount of white matter in the human brain [12]. Importantly, such anatomical differences have underlying molecular correlates as demonstrated by the atlas of the mid-gestational human brain [13].

Taking into account 65 million years of inter-species evolutionary divergence, it is not surprising that there are also significant discrepancies in both innate and adaptive immunity between human and mouse, including differences in immune receptors, cell types, and signaling pathways [14]. Such substantial inter-species differences can have considerable impact on drug discovery and development efforts. In fact, biomedical research has long relied on experimentation in mice to investigate human diseases and evaluate drug candidates. The value of animal models in drug discovery and development cannot be overstated even though the failure of the clinical trials can be attributed to other factors like poor design of the trials (wrong dose or endpoint), different genetic make-up among patients, and so many other logistic issues. However, the high rate of drug failures in general start right from selection of the correct molecule in pre-clinical studies and recent failures of AD therapies in phase III of clinical trials, in particular, again point to the fact that inter-species discrepancies at all biological levels should be seriously taken into consideration before proceeding to expensive clinical trials. Computational systems models can facilitate this task by gathering both experimental data and published knowledge, standardizing this information, integrating them across various biological scales, and representing this species-specific information in the form of consolidated cause-and-effect digital models. We have already shown the value of such approach for identification of disease-specific pathways in AD as compared to normal bioprocesses in the human brain [15], and Pappalardo et al. built computational model in immune system, which predicts how immune system activates in different conditions [16]. Motivated by these results, we sought to systematically model and mechanistically compare neuroinflammatory pathways specific to microglia, astrocytes, macrophages, and neurons between human and mouse in the context of AD. Biological Expression Language (BEL) [17] was used to build cause-and-effect computable models of neuroinflammation for both human and mouse based on published knowledge in the biomedical literature. Comparison of human and mouse models were performed at structural and functional levels, with the aim of answering the question, whether our current knowledge about neuroinflammation in mouse and human allows us to speak about a "functional equivalence" between these two species. In this work, we present the species-specific models and discuss which functional elements of neuroinflammation are similar and which elements are different between the two species. We also demonstrate, how the modeling approach can be used to explain—at a mechanistic level—the failure of translation from preclinical to clinical phase using a given drug in clinical phase III.

METHODS

Corpus selection and construction of neuroinflammation models specific for mouse and human

Based on the workflow illustrated in Fig. 1, we have built neuroinflammation BEL models for human and mouse by extracting knowledge from literature, which are specific to three cell types in neurons: astrocytes, microglia, and macrophages; as they are actively involved in neuroinflammation.

To build specific models for human and mouse, we have generated relevant literature corpora using SCAIView [18], a text mining tool developed in Fraunhofer SCAI. We specifically selected articles, which are specific to neuroinflammation in human as well as in mouse. We retrieved 189 documents for mouse using the following query: {*((((([NDD: "Neuroinflammation"]) AND [MeSH Disease: "Alzheimer Disease"]) AND [Organism:"Mus musculus"]) NOT [Organism:"Rattus norvegicus"])*

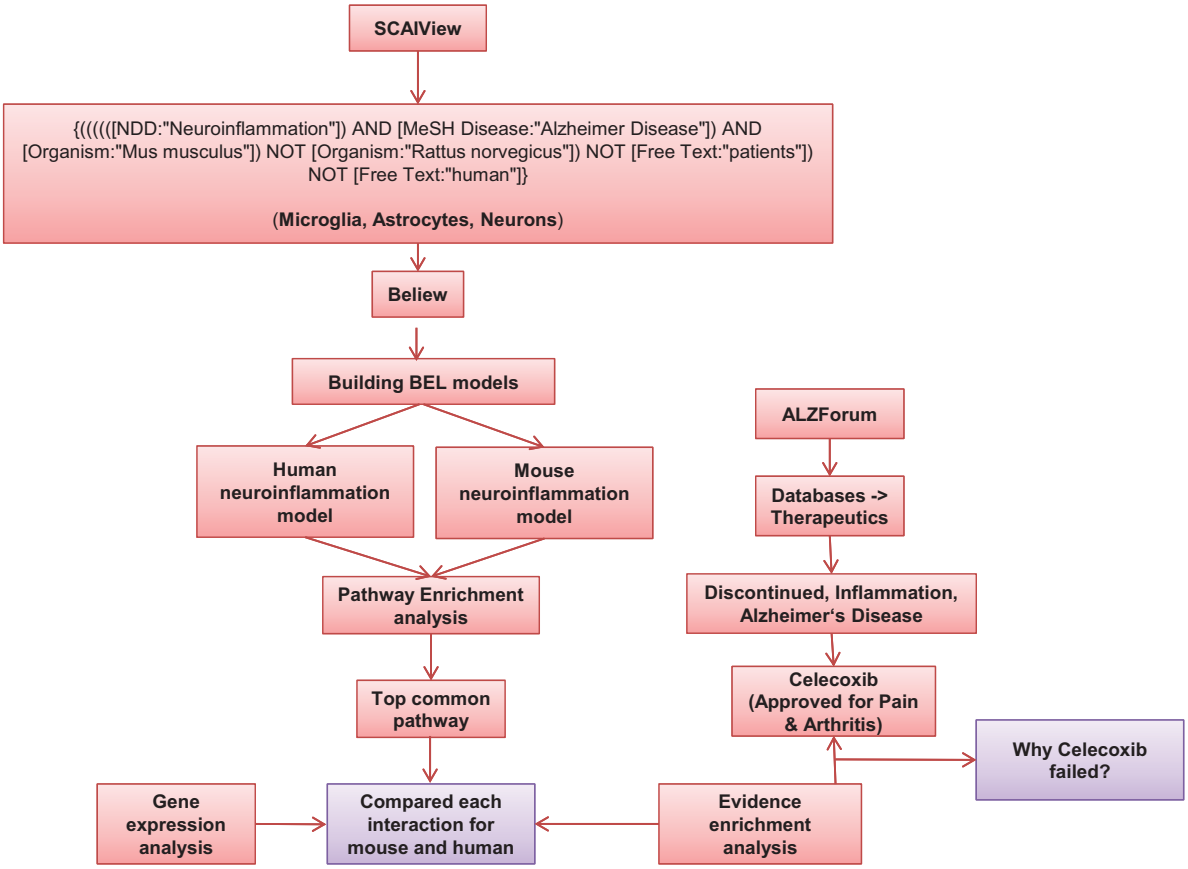


Fig. 1. Workflow used for construction of models and their analysis.

NOT [Free Text:"patients"]) NOT [Free Text:"human"]}. Out of 189 documents, we manually filtered 173 articles which were further incorporated into the 'Mouse neuroinflammation model'. Similarly, 309 articles were harvested for human using the query {*((((((((((NDD:"Neuroinflammation"]) AND [MeSH Disease:"Alzheimer Disease"]) AND [Organism:"Homo sapiens"]) NOT [Organism:"Mus musculus"]) NOT [Organism:"Rattus rattus"]) NOT [Free Text:"mice"]) NOT [Free Text:"mouse"]) NOT [Free Text:"murine"]) NOT [Organism:"Rattus sordidus"]) NOT [Free Text:"rat"]) NOT [Free Text:"rodent"]*} and 152 articles were used to build the 'Human neuroinflammation model'. The selected articles were subjected to analysis of causal and correlative relationships using the the BEL Information Extraction workflow (BELIEF) [19], a semi-automatic system that identifies cause-and-effect relationships in scientific text. The statements proposed by BELIEF were semi-automatically extracted, converted into BEL statements and further curated manually to build the neuroinflammation BEL models.

Comparison of mouse and human neuroinflammation models

Comparison based on interactions from species-specific BEL models

A systematic comparative analysis was done based on the molecular involvement of genes, bioprocesses, and pathways. The BEL models were used to compare pathways and we identified shared as well as unique pathways, based on BEL statements and entities. In addition to this, we have done a pathway enrichment analysis using DAVID for human and mouse, by giving the complete gene set as input from each of the models (Supplementary Table 1) and compared the most enriched pathways (Supplementary Table 2). To identify the consistency between mouse and human interactions, we have

done an additional manual evidence enrichment from the literature (Supplementary Table 3).

Comparison based on gene expression data for genes from species-specific models

To identify the concordance at the gene expression level between the two species, human and mouse, we have analyzed the representation of expressed genes in our models using gene expression datasets from Gene Expression Omnibus (GEO) [20]. We have analyzed 6 different gene expression datasets (Mouse: GSE35338, GSE74615, GSE74995 and Human: GSE26927, GSE45880, GSE59671) to support the findings of our study, all of which were related to neuroinflammation. GSE35338 contains expression data from astrocytes of mice where inflammation is induced by lipopolysaccharide treatment. Similarly, GSE74615 provides expression values from astroglia and microglia of transgenic mice, whereas GSE74995 has expression profiles of cortical tissue of AD transgenic mice. GSE45880 contains cytokine-induced expression profiles of human cerebral endothelial cell. GSE26927 contains expression data of males and females from different NDD patients, of which we considered only AD-related datasets. Lastly, GSE59671 contains expression values of RNAs of human smooth muscle cells treated with celecoxib and rofecoxib. All datasets were analyzed using the GEO2R tool provided by GEO [21].

RESULTS

Differential analysis of human and mouse neuroinflammatory pathways at the molecular level using cause-effect models

The neuroinflammation model for human consists of 884 BEL statements comprising 671 nodes and 1,224 edges extracted from 152 articles. Likewise, the mouse neuroinflammation model consists of 1,016 nodes and 1,939 edges specific to mouse, supported by 1,395 BEL statements from 173 articles. Even though we have a higher number of articles that discuss neuroinflammation in humans than mouse, we found that biological entities and relationships are in fact highly redundant among human specific articles. However, we could integrate more BEL statements and a larger variety of entities in the mouse model, as a higher number of novel molecular interactions have been studied in transgenic mouse experiments. For example, in case of App (Amyloid precursor protein), there are 155 transgenic mouse models, which have been generated for studies on amyloid biology [22].

In order to find shared and unique pathways between mouse and human, we have done a differential analysis between the two models. Gene set enrichment analysis was performed on genes in each model using the DAVID tool [23] to identify the most enriched pathways in both models. We retrieved 42 pathways in the human model and 29 pathways in the mouse model, of which 19 pathways were unique to human model and 24 pathways were found common between the two (Supplementary Table 2).

Among the 19 unique pathways in the human neuroinflammation model, we found VEGF signaling pathway and mTOR signaling pathway as the two top ranked pathways. In the case of mouse models, we found only 6 unique pathways but they were not specific to neuroinflammation. Based on these findings, we linked the bioprocesses corresponding to each pathway from the neuroinflammation models (Fig. 2). Despite shared pathways between the species, we found differential molecular patterns at the level of bioprocesses. For instance, bioprocesses like pyropotosis and pattern recognition receptor activity are better represented in the human model than the mouse model. However, when we extend these bioprocesses to pathways like cytokine-cytokine receptor interaction and Nod-like receptor signaling pathway, we can see that some parts of these pathways (e.g., inflammatory response or astrocyte activation) are well investigated in mouse experiments. Therefore, the resolution of mechanistic knowledge at the level of bioprocesses is higher in the mouse model than in the human model. At the abstraction level of canonical pathways, there are more commonalities between the two species, than at the level of underlying "cause-and-effect" mechanisms.

We also sought to identify to which extent the mouse model can represent human interactions at the molecular level. For this purpose, we investigated in more details the top common pathway (from DAVID analysis) between the two species; that is, cytokine-cytokine receptor interaction pathway. As shown in Fig. 3, there are 73 interactions in this pathway, which were represented in the human neuroinflammation model. Out of these 73 interactions, 33 interactions are protein-protein interactions at molecular level and 40 interactions among proteins, cell types, and bioprocesses, which are at cellular and bioprocess levels.

Furthermore, we have checked additional literature for more evidences on each interaction in the

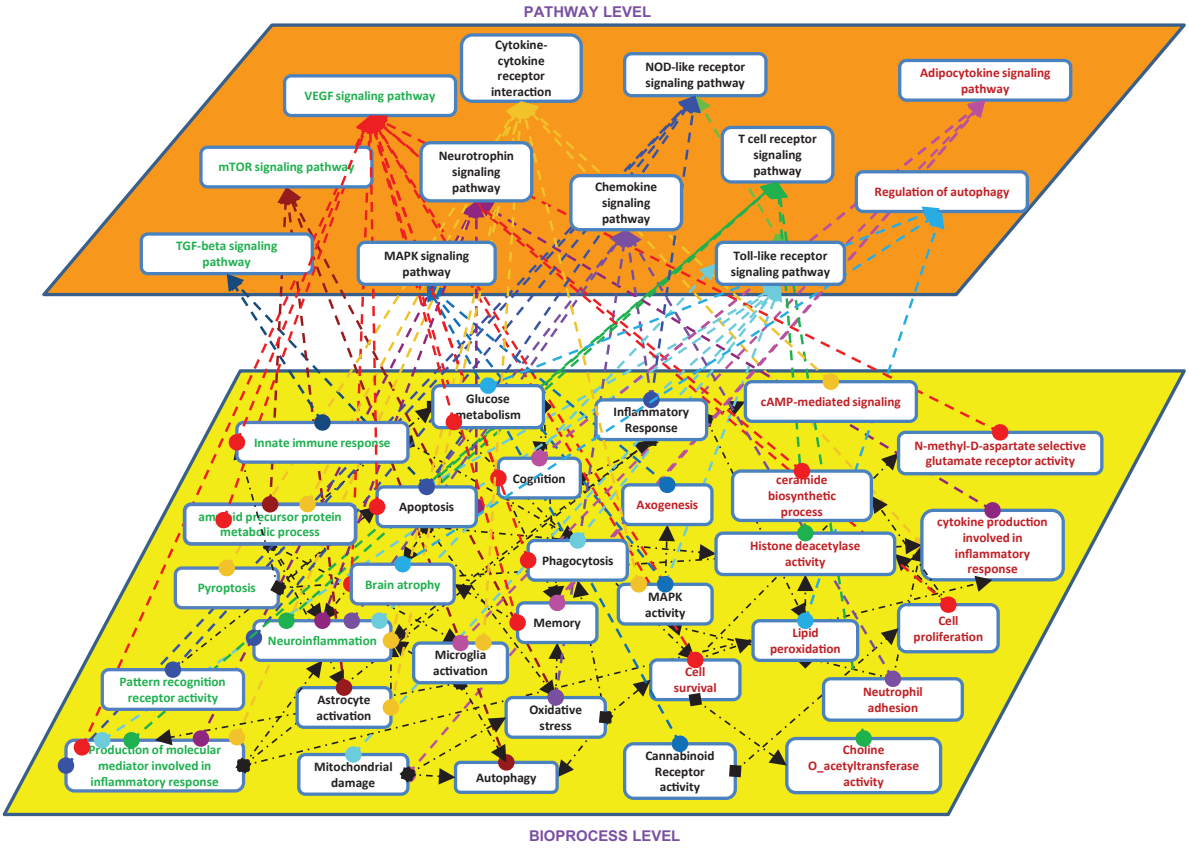


Fig. 2. Shared pathways and bioprocess between mouse and human. Entities present in both models are in black color, enriched pathways in mouse are in RED color, enriched pathways in human are in GREEN color.

human cytokine-cytokine pathway and compared these against the mouse model to see how many of the molecular and cellular interactions proven in humans are already reflected in the mouse model. At the molecular level, we found that 27% of cytokine interactions are similar in both species, 15% of interactions in mouse are found to be contradictory (opposite direction) to human, and 58% of interactions were found only in human or in other words, 58% of these human interactions are not proven with mouse experiments (Supplementary Table 4).

These numbers should be interpreted with regard to the bias of our knowledge repertoire toward molecular research and publication on mouse and human experiments. However, within these limits, it can be observed that our current knowledge in the domain of cytokine pathway not only reflects the contradictions in molecular interactions between human and mouse, but also misses many comparable human interactions in the mouse model.

At the cellular level, based on the interactions among cell types (microglia and astrocytes), bioprocesses, and proteins, we found a relatively higher similarity between the two species. 62% of the interactions were similar between the species, 28% of interactions were found only in human, and 10% of interactions were found to be contradicting in human and mouse.

In addition to the above comparison and to support our findings from model comparison, we have done an overall analysis on the availability of gene expression data for each species (Supplementary Table 5). We found 13 experiments with the topic of neuroinflammation in human and 32 experiments in mouse or rat (Supplementary Table 5). From these experiments, we have further selected GSE74615, GSE35338, and GSE74995 for mouse, as these experiments were performed using brain-specific tissues like astrocytes, microglia, and cortex. Based on these experimental data, we have investigated how many cytokine interactions in our model are supported by independent

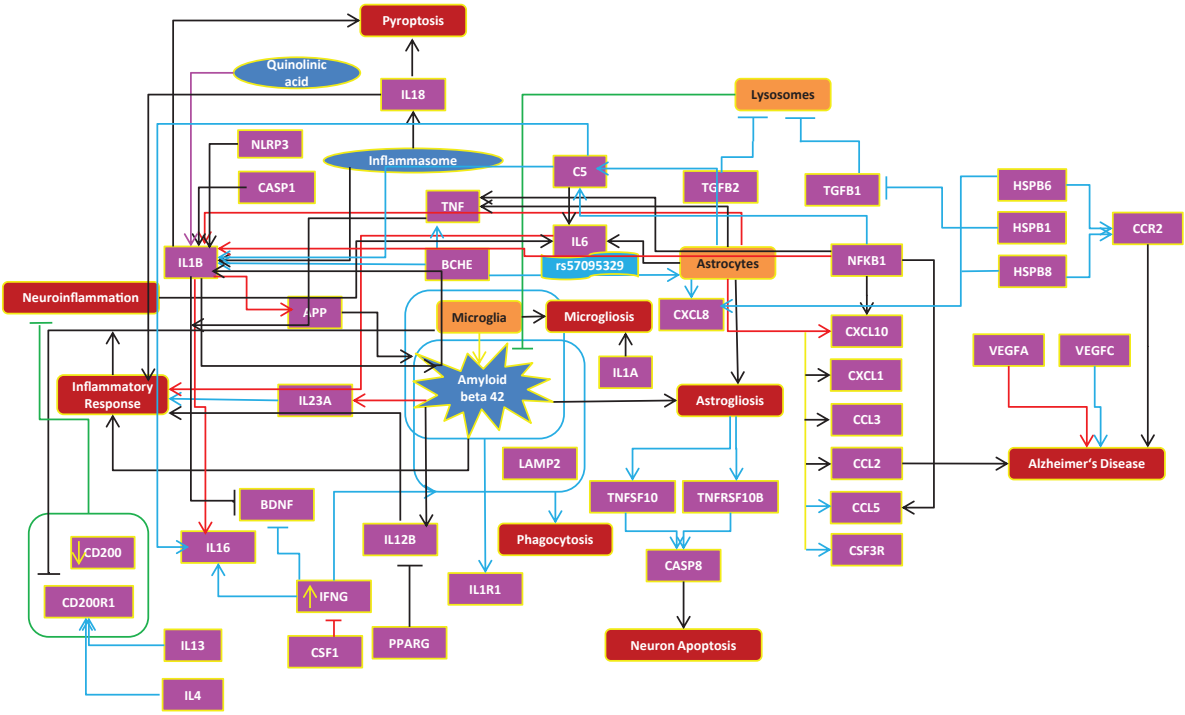


Fig. 3. Cytokine-cytokine receptor interaction (human and mouse). BLACK color indicates interactions are consistent with human, RED color depicts contradictory interactions in mouse compared to human, and BLUE arrows show the interactions found only in human.

transcriptome data. Similarly, we have conducted gene expression analysis for human brain tissues using GSE26927 (entorhinal cortex) and GSE45880 (cerebral endothelial cell line) datasets. Since these experiments have not used the same tissues in human and mouse, a direct comparison between the two species is not possible, considering the fact that genes can be expressed differently in different tissues and regions. Excluding tissue specificity, expression patterns for 31% of genes in both species were the same, while 14% of genes from Fig. 3 were significantly expressed only in human and 19% of genes were expressed only in mouse. 17% of genes were identified to be inconsistent (same gene is shown as up- and downregulated) within the species and 7% of genes showed to have contradictions between the species. 12% of genes were found to be statistically non-significant (p-value >0.05) in both species (Supplementary Table 6).

Analysis of failure of drugs on the basis of translation between species

Based on the approach suggested by Younesi and Hofmann-Apitius for translational validation of disease models [24], we aimed to identify the extent to which the mouse model can translate into the human biological interactions by including the mode-of-action of drugs within the mechanistic model. For this purpose, we performed an analysis based on failed drugs in AD that specifically were targeted against neuroinflammation as these drugs proved to work in mouse models in pre-clinical development, but failed in human during clinical trials.

Accordingly, drugs which have failed in AD clinical trials were collected from the Therapeutics database of AlzForum [25], using the following search query: Food and Drug Administration (FDA) status – discontinued, Target Type – Inflammation, Therapy Type – All, Condition – Alzheimer's disease and Mild cognitive impairment.

We were able to retrieve 8 failed or discontinued AD therapeutics, out of which, celecoxib which is approved for pain and arthritis was selected for analysis [26, 27]. The rationale behind this selection is that there were many lines of evidence supporting the role of comorbidity association between rheumatoid arthritis and AD [28–30], and points to a likely shared mechanism at the molecular and cellular levels. Thus, we performed mechanistic analysis around the targets of celecoxib, both for human and mouse, in order to find probable mechanistic differences in the

translation of interactions between the two species which might have led to the failure of celecoxib in AD.

Celecoxib has two main targets namely: PTGS2 (prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)) and PDPK1 (3-phosphoinositide dependent protein kinase 1).

According to transgenic mouse experiments, in normal conditions Pdpk1 increases the activity of Il4 and Ins [31] and also phosphorylates Gsk3b and Akt1 [32]. Pdpk1 also inhibits Ccr2, M1 macrophages, and insulin resistance [31]. Similarly, Ptgs2 increases Aβ peptides [33, 34], which further increases Tnf and Nfkb1 [35]. Nfkb1 increases Ptgs2 forming a self-regulatory network [36] leading to an increase in Aβ peptide aggregation and an increase in inflammation due to increased activation of Tnf. Based on the above-mentioned interactions deduced from mouse experiments, inhibition of PTGS2 and PDPK1 with administration of celecoxib seems to be a good tactic in treating neuroinflammation. Here are some of the positive effects of celecoxib in case of AD and neuroinflammation (mainly based on mouse experiments and few supportive evidence from human experiments):

- Celecoxib increases M1-macrophage[1] and Ccr2 and thereby increases phagocytosis and Aβ clearance in mouse models respectively [31].
- Furthermore, Pdpk1 inhibition of Gsk3b phosphorylation by celecoxib prevents the formation of neurofibrillary tangles through phosphorylated tau (Mapt).

However, we have extended our investigation very specific to the celecoxib interactions particularly in the context of AD on the basis of mouse models.

We have also deduced the perturbation caused in normal physiological brain pathway in human upon administration of celecoxib. The following lines of evidence provide, at mechanism level, explanatory insight why celecoxib could not work in humans as expected in mouse:

- In the case of Pdpk1 inhibition, phosphorylation of Akt1 can be reduced which may further increase the phosphorylation of Tsc2. According to Shang et al., it was proposed that phosphorylation of threonine at position 1462 of Tsc2, a target of Akt1, is increased in AD [37], and supported by the finding that Tuberin (TSC2) was hyperphosphorylated at Thr1462 in postmortem frontal cortex tissue of both AD and PD/DLB patients [38].
- Hyperphosphorylated Tsc2 hyperactivates Mtor through Rheb which reduces autophagy [39–41]. If autophagy is reduced, then it will lead to increased amyloid deposition.
- Also, a very recent paper by Oddo et al. stated that decreased mTOR activity may be necessary to decrease BACE1 and reduce Aβ generation in AD from mouse experiments [41]. Therefore, as a result of reduction in Akt1 activity upon celecoxib administration, Mtor hyperactivates and leads to increased amyloid deposition.
- Celecoxib increases IL-4 (which is anti-inflammatory protein) inhibition which causes inflammation [43]. Similarly, inhibition of Pdpk1 by celecoxib might cause increase in insulin resistance through inhibition of Ins. Insulin resistance is proposed to be a risk factor for AD [44].
- Upon celecoxib administration, it inhibits PDPK1 (PDK1) which reduces the phosphorylation of AMPK (PRKAA1) (which in normal physiology reduces MTOR activation) leading to the inhibition of autophagy. This leads to decrease in Aβ clearance [45, 46]. This disease mechanism has been described already in detail by Kodamullil et al. [15].

Based on the above reports (evidence which support the usage of celecoxib with mouse experiments and evidence how the normal physiological mechanism in human brain is perturbed upon administration of celecoxib), we can conclude that even though celecoxib modulates MTOR toward neuronal protection to limit the toxicity of Aβ and consequently neuroinflammation in AD, we may also require targeting TSC2, AKT, and AMPK simultaneously. It is noteworthy at this point that AD and neuroinflammation in humans are so complicated that it appears unlikely that an experimental design based on a specific transgenic mouse manipulated for a single gene allows us to expose all the interlinked mechanisms.

To validate the celecoxib interactions shown in Fig. 4, we have done gene expression analysis using the gene expression experiment GSE59671, which

[1]The authors are aware of the fact that M1/ M2 nomenclature of macrophages is most likely obsolete or needs new interpretation as pointed out by Xue J. et al. 2014 (Immunity. 2014 Feb 20; 40(2): 274–288), but we stick to the "M1 object" in our models, as defined processes and properties have been associated with this object are derived from other researchers.

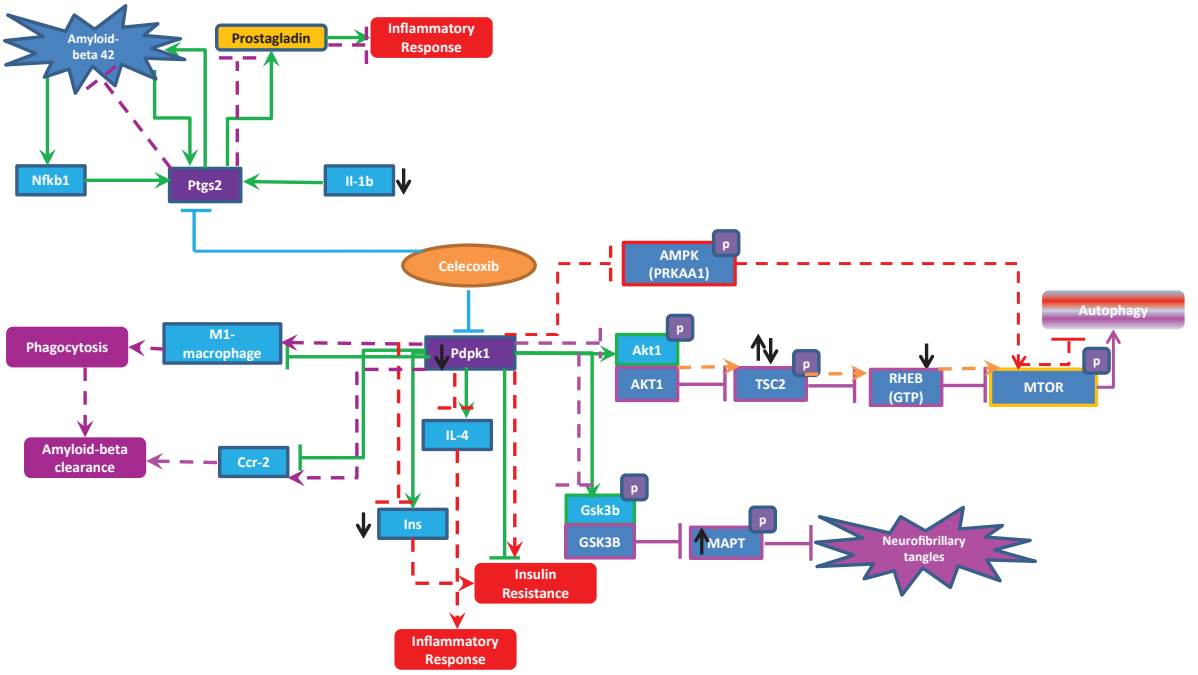


Fig. 4. Celecoxib interactions in human and mouse. The green lines represent the normal brain interaction of celecoxib targets (Pdpk1 and Ptgs2) with other entities proposed in mouse experiments. The red dotted lines indicate how normal pathways might be perturbed upon administration of celecoxib, which could lead to severity of neuroinflammation and AD. The purple dotted lines are the beneficial effects of celecoxib in the context of AD and neuroinflammation. The black up and down arrows represents the expression of genes (upregulation and downregulation) upon celecoxib administration.

is the only dataset available related to our research context, even though the cell line used was human aortic smooth muscle cells (3F1243) pre-treated with rofecoxib (500 nM) or celecoxib (500 nM) (Supplementary Table 5). As seen in Fig. 4, the expression data supports the reduction in levels of INS, Il1b, and RHEB, and increase in TSC2 although there are inconsistencies about increased MAPT in human upon celecoxib administration.

DISCUSSION

Mouse models are extensively used in biomedical research mainly to understand the etiology of the disease. Complex diseases like AD may involve several simultaneous alterations in molecular and processual activities, including neuroinflammation, aggregation of Aβ peptides, or tau phosphorylation, which are likely to contribute to pathophysiology. In this paper, we have compared the mouse and human at molecular, cellular, and pathway levels to shed light on mechanistic differences with important implications for translation outcomes. Mechanistic modelling specific to species allows us to "embed" and "represent" similarities and differences in innate immunity which can lead to the development of "conflictious information detection engine". It is important to note that our analysis is purely based on the research and publication bias in mouse and human experiments as many mouse experiments are mainly focused on particular explorative areas, and experiments with human tissues are also concentrated on limited areas of disease mechanism. We found that mouse experiments often reveal new molecular interactions between different entities that are not observed or reported in human experiments. Differential analysis of mouse and human model for neuroinflammation shows that mouse and human differ at the molecular and cellular levels, but have more similarities at the pathway levels as numbers indicate. More explicitly, the underlying molecular patterns which lead to a particular bioprocess differ between the two species. This finding implies that although the two species share some similarities at the cellular or pathway level, the pattern of molecular interactions that form, govern, and regulate those pathways is substantially different between mouse and human.

It is notable that mouse models have provided significant insights into many disease areas like cancer;

acute promyelocytic leukemia. However, recent drug failures in the area of neurodegeneration have put a question mark behind the extent to which mouse models have been used in preclinical drug discovery and to what extent transgenic mice mimic human brain pathophysiology mechanisms. Pathophysiology mechanisms are likely to act together and they seem to be organized in a temporal cascade of events that ultimately result in a severe disease phenotype. Experiments with single gene knock-out in mice can reveal only minor aspects of the disease perturbations and do not usually allow us to decipher the full complexity of the mechanisms underlying the disease. For example, even though high amounts of Aβ are observed in APP knock-in mice carrying the Swedish mutations, these mice do not produce amyloid plaques [47]. On the other hand, human APP K670N-M671L (APPSw), which have amyloid deposition and behavioral deficits, do not exhibit any neuronal loss [48]. This points to the fact that each strain of mouse results in various phenotypes and do not represent the main clinical outcome. This emphasizes the need to do systematic comparisons between the model organism (and factual findings in mice and rats) and human. Additionally, development of various mouse models should also consider the absence of key functional human genotypes (Apoe 3,4) in animal models. If the above hypothesis regarding systematic differences at molecular level among species holds true, then the expectation is to observe different or multipoint translational outcomes in human compared to mouse. The most striking case of a different outcome happens when a drug fails and the most common case of a multipoint outcome is serendipitous effect of a drug on an unexpected biology. Even in the case that drug candidate successfully hits the pathology, the subsequent side effects clearly show the underlying mechanistic differences between human and mouse. Comparative analysis of the mode-of-action of celecoxib in the neuroinflammatory pathway between human and mouse at the high-resolution molecular level demonstrates that perhaps target studies ignore human unique pathways and the underlying unique mechanisms. It was found that many off-target interactions that could occur in human were not considered in the mouse experiments. The fact that mimicking human disease pathology in mice using a chemical agent or a single gene is purely correlative and supports the notion that it is crucial to take the fundamental mechanistic differences between mouse and human into consideration when attempting to translate preclinical findings to clinical trials. This is not intended to criticize the use of mouse models (considering the fact that failure of clinical trials are not solely associated with mouse models, rather also to differences in patient level, drug dosage, etc.) but rather to point out the repetitive failure of clinical trials in AD and neuroinflammation indications. Therefore, it is time to rethink about the caveats inherent with the mouse model experiments. The construction and simulation of computable cause-and-effect models of disease pathology can greatly increase the probability of translational success. The computable cause-and-effect modeling approach described in this work can be complemented with a systems biology simulation at systems level. Such in-silico models can effectively contribute to well-informed design of *in vivo* mouse models by predicting the expected and unexpected outcomes compared to human conditions. We foresee that with the advent of big biomedical data and growth of published knowledge, disease-specific computable models will play an important role in drug discovery and biomarker identification for clinical applications.

ACKNOWLEDGMENTS

The authors would like to thank Life Science Informatics batch 2015-2016 students who participated in curation works of BEL statements. The authors also would like to acknowledge the support from the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115568 (project AETIONOMY), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2017-2013) and European Federation of Pharmaceutical Industries and Associations companies' in kind contribution.

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/17-0255r2).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-170255.

REFERENCES

[1] Amor S, Puentes F, Baker D, Van Der Valk P (2010) Inflammation in neurodegenerative diseases. *Immunology* **129**, 154-169.

[2] Heneka MT, Carson MJ, El Khoury J, Landreth GE, Brosseron F, Feinstein DL, Jacobs AH, Wyss-Coray T, Vitorica J, Ransohoff RM, Herrup K, Frautschy SA, Finsen B, Brown GC, Verkhratsky A, Yamanaka K, Koistinaho J, Latz E, Halle A, Petzold GC, Town T, Morgan D, Shinohara ML, Perry VH, Holmes C, Bazan NG, Brooks DJ, Hunot S, Joseph B, Deigendesch N, Garaschuk O, Boddeke E, Dinarello CA, Breitner JC, Cole GM, Golenbock DT, Kummer MP (2015) Neuroinflammation in Alzheimer's disease. *Lancet Neurol* **14**, 388-405.

[3] Heppner FL, Ransohoff RM, Becher B (2015) Immune attack: The role of inflammation in Alzheimer disease. *Nat Rev Neurosci* **16**, 358-372.

[4] Pimplikar, Sanjay W (2014) Neuroinflammation in Alzheimer's disease: From pathogenesis to a therapeutic target. *J Clin Immunol* **34**, 64-69.

[5] Heneka MT, Kummer MP, Latz E (2014) Innate immune activation in neurodegenerative disease. *Nat Rev Immunol* **14**, 463-77.

[6] Birch AM, Katsouri L, Sastre M (2014) Modulation of inflammation in transgenic models of Alzheimer's disease. *J Neuroinflammation* **11**, 25-25.

[7] Nazem A, Sankowski R, Bacher M, Al-Abed Y (2015) Rodent models of neuroinflammation for Alzheimer's disease. *J Neuroinflammation* **17**, 74-74.

[8] Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, Richards DR, McDonald-Smith GP, Gao H, Hennessy L, Finnerty CC (2013) Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A* **110**, 3507-12.

[9] Van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, Macleod MR (2010) Can animal models of disease reliably inform human studies? *PLoS Med* **7**, e1000245.

[10] Takao K, Miyakawa T (2015) Genomic responses in mouse models greatly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A* **112**, 1167-1172.

[11] Warren HS, Tompkins RG, Moldawer LL, Seok J, Xu W, Mindrinos MN, Maier RV, Xiao W, Davis RW (2015) Mice are not men. *Proc Natl Acad Sci U S A* **112**, E345-E345.

[12] Treuting PM, Dintzis SM (2011) Comparative Anatomy and Histology: A Mouse and Human Atlas (Expert Consult). Academic Press.

[13] Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, Arnold JM (2014) Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206.

[14] Mestas J, Hughes CC (2004) Of mice and not men: Differences between mouse and human immunology. *J Immunol* **172**, 2731-2738.

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

[16] Pappalardo F, Fichera E, Paparone N, Lombardo A, Pennisi M, Russo G, Leotta M, Pappalardo F, Pedretti A, De Fiore F, Motta S (2016) A computational model to predict the immune system activation by citrus-derived vaccine adjuvants. *Bioinformatics* **32**, 2672-2680.

[17] OpenBEL Framework and language: http://openbel.org/

[18] Text-mining tool SCAIView: http://www.scaiview.com

[19] Semi-automatic BEL extraction tool, BELIEF: http://belief.scai.fraunhofer.de/BeliefDashboard/

[20] GEO database: https://www.ncbi.nlm.nih.gov/geo/

[21] GEO2R, gene expression analysis tool: http://www.ncbi.nlm.nih.gov/geo/geo2r/

[22] Mouse models for Alzheimer's Disease in MGI database: http://www.informatics.jax.org/disease/104300 last accessed on November 29, 2016.

[23] Gene functional annotation tool, DAVID: https://david.ncifcrf.gov/

[24] Younesi E, Hofmann-Apitius M (2013) From integrative disease modeling to predictive, preventive, personalized and participatory (P4) medicine. *EPMA J* **4**, 23.

[25] AlzForum for discontinued drugs in Alzheimer's: http://www.alzforum.org/therapeutics

[26] Shishodia S, Koul D, Aggarwal BB (2004) Cyclooxygenase (COX)-2 inhibitor celecoxib abrogates TNF-induced NF-κB activation through inhibition of activation of IκBα kinase and Akt in human non-small cell lung carcinoma: correlation with suppression of COX-2 synthesis. *J Immunol* **173**, 2011-2022.

[27] Chun KS, Kim SH, Song YS, Surh YJ (2004) Celecoxib inhibits phorbol ester-induced expression of COX-2 and activation of AP-1 and p38 MAP kinase in mouse skin. *Carcinogenesis* **25**, 713-722.

[28] Ferraccioli G, Carbonella A, Gremese E, Alivernini S (2012) Rheumatoid arthritis and Alzheimer's disease: genetic and epigenetic links in inflammatory regulation. *Discov Med* **14**, 379-388.

[29] Jenkinson ML, Bliss MR, Brain AT, Scott DL (1989) Rheumatold arthritis and senile dementia of the Alzheimer's type. *Rheumatology* **28**, 86-87.

[30] Beard CM, Kokman E, Kurland L (1991) Rheumatoid arthritis and susceptibility to Alzheimer's disease. *Lancet* **337**, 1426.

[31] Kawano Y, Nakae J, Watanabe N, Fujisaka S, Iskandar K, Sekioka R, Hayashi Y, Tobe K, Kasuga M, Noda T, Yoshimura A (2012) Loss of Pdk1-Foxo1 signaling in myeloid cells predisposes to adipose tissue inflammation and insulin resistance. *Diabetes* **61**, 1935-1948.

[32] Pollock CB, Yin Y, Yuan H, Zeng X, King S, Li X, Kopelovich L, Albanese C, Glazer RI (2011) PPARδ activation acts cooperatively with 3-phosphoinositide-dependent protein kinase-1 to enhance mammary tumorigenesis. *PLoS One* **6**, e16215.

[33] Cakała M, Malik AR, Strosznajder JB (2007) Inhibitor of cyclooxygenase-2 protects against amyloid β peptide-evoked memory impairment in mice. *Pharmacol Rep* **59**, 164-172.

[34] Ferretti MT, Bruno MA, Ducatenzeiler A, Klein WL, Cuello AC (2012) Intracellular Aβ-oligomers and early inflammation in a model of Alzheimer's disease. *Neurobiol Aging* **33**, 1329-1342.

[35] Wang HM, Zhao YX, Zhang S, Liu GD, Kang WY, Tang HD, Ding JQ, Chen SD (2010) PPARγ agonist curcumin reduces the amyloid-β-stimulated inflammatory responses in primary astrocytes. *J Alzheimers Dis* **20**, 1189-1199.

[36] Zhang X, Luhrs KJ, Ryff KA, Malik WT, Driscoll MJ, Culver B (2009) Suppression of nuclear factor kappa B ameliorates astrogliosis but not amyloid burden in APPswe/PS1dE9 mice. *Neuroscience* **161**, 53-58.

[37] Chen Shang Y, Zhong Chong Z, Wang S, Maiese K (2013) Tuberous sclerosis protein 2 (TSC2) modulates CCN4 cytoprotection during apoptotic amyloid toxicity in microglia. *Curr Neurovasc Res* **10**, 29-38.

[38] Habib SL, Michel D, Masliah E, Thomas B, Ko HS, Dawson TM, Abboud H, Clark RA, Imam SZ (2008) Role of tuberin in neuronal degeneration. *Neurochem Res* **33**, 1113-1116.

[39] Shahani N, Pryor W, Swarnkar S, Kholodilov N, Thinakaran G, Burke RE, Subramaniam S (2014) Rheb GTPase regulates β-secretase levels and amyloid β generation. *J Biol Chem* **289**, 5799-5808.

[40] Jacobson O, Chen X (2013) Interrogating tumor metabolism and tumor microenvironments using molecular positron emission tomography imaging. Theranostic approaches to improve therapeutics. *Pharmacol Rev* **65**, 1214-1256.

[41] Saxton RA, Sabatini DM (2017) mTOR signaling in growth, metabolism, and disease. *Cell* **168**, 960-976.

[42] Velazquez R, Shaw DM, Caccamo A, Oddo S (2016) Pim1 inhibition as a novel therapeutic strategy for Alzheimer's disease. *Mol Neurodegener* **11**, 52-52.

[43] Woodward EA, Prêle CM, Nicholson SE, Kolesnik TB, Hart PH (2010) The anti-inflammatory effects of interleukin-4 are not mediated by suppressor of cytokine signalling-1 (SOCS1). *Immunology* **131**, 118-127.

[44] Willette AA, Bendlin BB, Starks EJ, Birdsill AC, Johnson SC, Christian BT, Okonkwo OC, La Rue A, Hermann BP, Koscik RL, Jonaitis EM (2015) Association of insulin resistance with cerebral glucose uptake in late middle–aged adults at risk for Alzheimer disease. *JAMA Neurol* **72**, 1013-1020.

[45] Williams MR, Arthur JSC, Balendran A, van der Kaay J, Poli V, Cohen P, Alessi DR (2000) The role of 3-phosphoinositide-dependent protein kinase 1 in activating AGC kinases defined in embryonic stem cells. *Curr Biol* **10**, 439-448.

[46] Mora A, Komander D, van Aalten DM, Alessi DR (2004) PDK1, the master regulator of AGC kinase signal transduction. *Semin Cell Dev Biol* **15**, 161-170.

[47] Elder GA, Gama Sosa MA, De Gasperi R (2010) Transgenic mouse models of Alzheimer's disease. *Mt Sinai J Med* **77**, 69-81.

[48] Irizarry MC, McNamara M, Fedorchak K, Hsiao K, Hyman BT (1997) APPSw transgenic mice develop age-related Aβ deposits and neuropil abnormalities, but no neuronal loss in CA1. *J Neuropathol Exp Neurol* **56**, 965-973.

Summary

Animal models were a significant contribution in efforts to decipher the causes of Alzheimer's disease. However, repeated failures in AD drug development motivated me to research on the predictability of mouse models in preclinical drug discovery and investigate the extent to which transgenic mice mimic human brain pathophysiological mechanisms. As outlined in the chapter 2 publication, mechanistic modelling specific to species allows us to "embed" and "represent" similarities and differences in the context of neuroinflammation. We found that mouse experiments often reveal molecular interactions between entities that are not observed or reported in human experiments. Though differential analysis of mouse and human models of neuroinflammation show that mice and humans differ at the molecular and cellular levels, they have similarities at the pathway level.

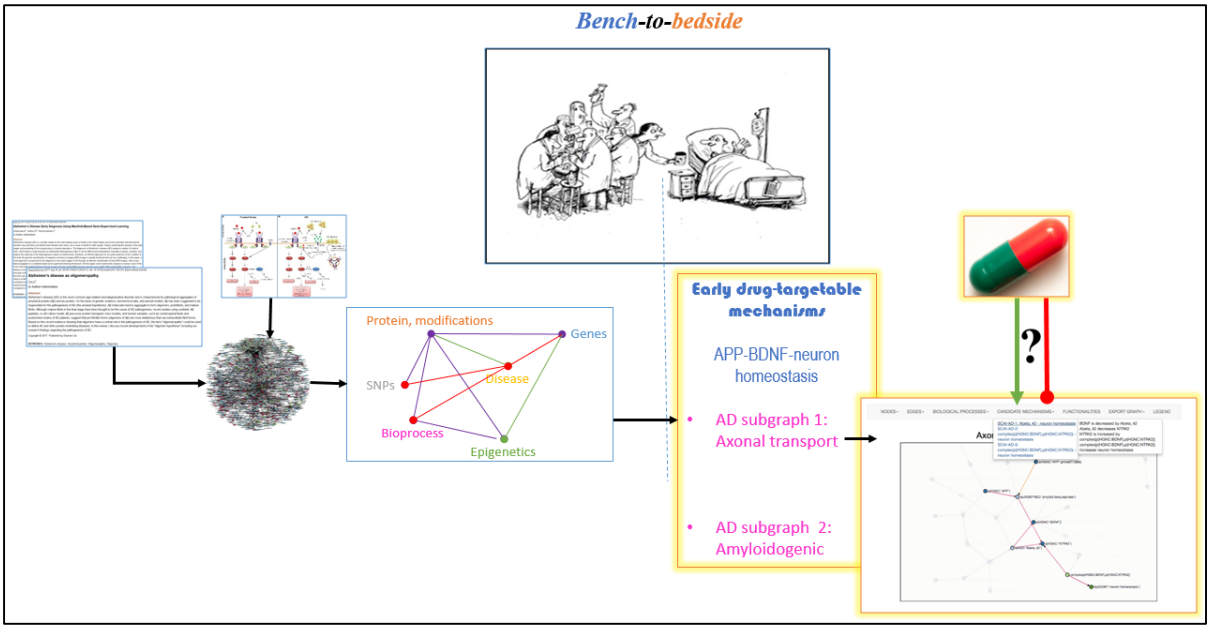
I also analyzed, how a discontinued drug, Celecoxib, failed based on the underlying mechanistic differences between human and mouse. Comparative analysis of the mode-of-action of Celecoxib in the neuro-inflammation demonstrates that perhaps target studies ignore pathways unique to humans and their underlying unique mechanisms. It is time then to rethink about how "translational" mouse experiments are and consider computational models as useful predictors of drug targets and side effects. Simulations of cause-and-effect models to elucidate mechanisms of action of drugs can greatly increase the probability of translational success at the pre-clinical stage.

It is important to note that our analysis was based purely on the research in mice and humans and is therefore subject to publication bias. Notably, many mouse experiments are focused on

explorative mechanisms, and experiments in human tissues concentrate on limited areas of disease mechanism.

Chapter 3

Mapping investment to possible drug target mechanisms



Introduction

Based on the mechanistic models introduced in Chapter 1 and 2, we have developed a knowledge base called NeuroMMSig (see additional publications: Domingo-Fernández D, Kodamullil AT et al. (2017)) representing essential pathophysiological mechanisms of neurodegenerative diseases. In the case of AD, we have identified more than 100 disease mechanisms which could potentially be targeted by drugs. In this chapter, I discuss the triangulation of "the landscape of mechanisms, targets and R&D investments" and establish, how wide the spectrum of possible candidate mechanisms to be targeted could potentially be. I do also trace the money that actually goes into a very limited number of well-established disease mechanisms. The main aim of my analysis is to highlight and prioritize the vast space of opportunities that exist outside of the "classical paradigm / pathways" of drug discovery

in AD. It was and is my explicit goal, to foster discourse about alternatives to the established, exclusively targeted mechanisms.

BIOBUSINESS BRIEFS

TRIAL WATCH

Tracing investment in drug development for Alzheimer disease

The high failure rates of drug R&D for Alzheimer disease (AD) — and particularly the recent failures of several drugs that target the amyloid cascade in phase III trials — have raised questions about the relative emphasis on particular therapeutic strategies for AD. Here, we present the results of an analysis that aims to inform such debate by mapping R&D investment to the pathophysiological mechanisms of AD, rather than to individual drugs or targets (see Supplementary information S1 (box) for details). The analysis exploits a 'mechanism inventory' for AD, based on a mechanistic model of the disease encoded in biological expression language (*Alzheimers Dement.* **11**, 1329–1339; 2015). This inventory, which has recently been developed (and continues to be enhanced) in the course of the AETIONOMY project (see Further information), includes 126 mechanisms for AD (*Bioinformatics* btx399, 23 June 2017).

A list of drugs that either underwent or were still undergoing clinical trials for AD was obtained from the Alzforum database in March 2016 (see Further information). We identified 59 discontinued drugs, 88 drugs in different phases of clinical trials that have not yet been discontinued and 5 approved drugs. Specific information regarding the drug targets of small molecules was extracted using DrugBank in combination with the HGNC and UniProt databases. This enabled each drug to be mapped to the most prominently targeted mechanism in the inventory. All antibodies that have been tested target either the amyloid cascade or tau protein, and were classified accordingly. A substantial proportion (36%) of drugs could not be mapped to a particular mechanism, and were classified as 'other'.

Investments in the development of these drugs for AD were estimated based on both the opinions of experts from various companies on clinical trial costs and on publicly available information on completed clinical trials for potential AD therapies conducted up until March 2016 (Supplementary information S1 (box)). ClinicalTrials.gov was used to gather information about the phase and number of patients enrolled in each trial. We also made a substantial effort to fill in missing information (see Supplementary information S2 (table) for an example). It is likely that many clinical trials have not been registered on ClinicalTrials.gov (particularly earlier in our study period, before registration became standard practice) and so we had to make assumptions about average trial sizes and associated cost per patient for some of the drugs.

Linking the investments in the clinical trials of each drug to their mapped mechanism provides an estimate of the relative amount of investment in each mechanism. FIG. 1a shows the estimated relative amounts of unsuccessful investment in different mechanisms, based on 37 of the 59 drugs that have been discontinued for which we could identify sufficient information for analysis (see Supplementary information S3 (table)). This analysis indicates that the mechanisms that were the greatest focus of R&D investment were the amyloid cascade, tau aggregation, neuroinflammation and neurotransmission. FIG. 1b shows the focus of current investment in different mechanisms based on 61 of the 88 drugs in ongoing clinical trials for which mechanisms could be assigned (see Supplementary information S4 (table)), indicating relatively little or no investment in some promising novel mechanisms such as targeting endocytosis or autophagy.

The main limitation of our study was the limited access to (or missing) data, which required assumptions to be made on the overall cost of trials. The estimates of cost also ignored inflation and changes in currency exchange rates over time. Overall, our intention is to make R&D investments in different mechanisms of AD more transparent. We hope that our work stimulates and informs the discussion about 'where the money goes' and catalyses the exploration of novel AD mechanisms.

Alpha Tom Kodamullil is at the Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin 53754, Germany.

Firas Zekri is at the University of Sfax, B.P. 1088, 3018, Tunisia.

Meemansa Sood is at the Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin 53754, Germany.

Bastian Hengerer is at Boehringer Ingelheim Pharma, 88397 Biberach an der Riss, Germany.

Luc Canard is at Sanofi, Chilly-Mazarin Cedex 91385, France.

Duncan McHale was previously at UCB Pharma, Belgium. Present address: Weatherden Ltd, Canterbury, Kent CT3 1TB, UK.

Martin Hofmann-Apitius is at the Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin 53754, Germany.

Correspondence to M.H-A.
martin.hofmann-apitius@scai.fraunhofer.de

doi:10.1038/nrd.2017.169
Published online 23 Oct 2017

Acknowledgements
The authors acknowledge support from the Innovative Medicines Initiative grant 115568 (AETIONOMY), and thank R. Karki and C. Ebeling for their support in installation of the MySQL database and queries.

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
AETIONOMY project: http://www.aetionomy.eu
Alzforum database: http://www.alzforum.org/therapeutics

SUPPLEMENTARY INFORMATION
See online article: S1 (box) | S2 (table) | S3 (table) | S4 (table)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF

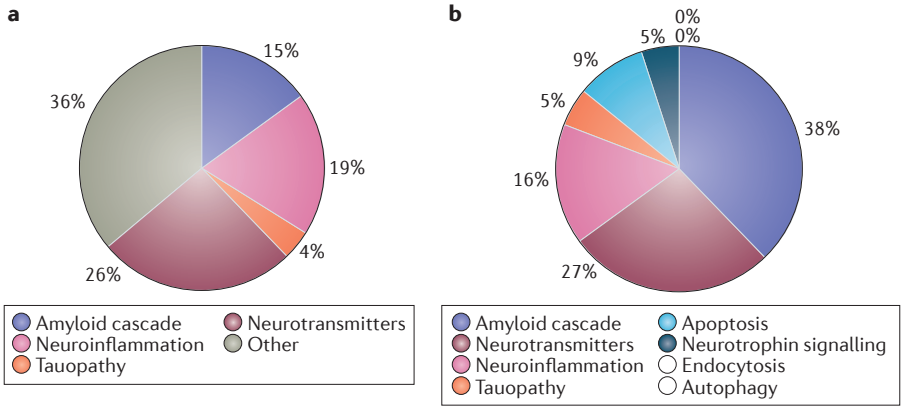


Figure 1 | **Investment in drug development for Alzheimer disease. a** | Overview of unsuccessful investment in mechanistic classes based on discontinued drugs. **b** | Mechanistic focus of investment in ongoing clinical trials. See Supplementary information S1 (box) for details.

In format provided by Hofmann *et al.* (doi:10.1038/nrd.2017.169)

Box 1 | **Data sources and analysis**

Estimating average clinical trial costs for drugs for Alzheimer disease

Estimating the average cost for the clinical development of drugs for Alzheimer disease (AD) is non-trivial. Our cost estimate is based on expert opinions from leading scientists in the area of neurology, psychiatry and computational biology working in various pharmaceutical companies. We also took into account and used publicly available information from scientific reports (Supplementary information S2 (table)). Based on these information sources, we calculated the average cost of an AD drug in each phase for each patient as shown in Table 1. We invite experts from other pharma and biotech companies to share their cost estimates with us, in order to broaden the evidence base for our current estimates.

Table 1 | Cost per patient per trial phase

Phase	Cost per patient*	Average cost used for calculation in this study
Phase I	80,000-100,000 Euros	100,000 Euros
Phase II	100,000-120,000 Euros	120,000 Euros
Phase III	60,000-80,000 Euros	80,000 Euros
Phase I+II	80,000-100,000 + 100,000-120,000 Euros	100,000 + 120000 = 220,000 Euros
Phase II+III	100,000-120,000 + 60,000-80,000 Euros	120,000 + 80,000= 200,000 Euros
Phase IV	US$2,992	2,640 Euros
Phase 0	US$500,000	441,746 Euros

*Based on pharma experts and/or scientific reports.

The main limitation of our study was the limited access to data or missing data, which required assumptions to be made on the overall cost of trials. The estimates of cost also ignored inflation and changes in currency exchange rates over time.

Identifying AD drugs tested in clinical trials

ClinicalTrials.gov (https://clinicaltrials.gov/) was used to retrieve information in March 2016 about the phase of the trial and the number of patients enrolled in trials for AD. In total, there were 1,536 AD clinical trials conducted between April 1995 **and** September 2014**,** of which 1,186 trials were aimed at treatment, prevention, diagnosis or supportive care for AD.

To generate an overview on drugs that either underwent or are still undergoing clinical trials we used the Alzforum database (http://www.alzforum.org/therapeutics) as an information source in March 2016. We retrieved 59 discontinued drugs, 88 drugs in different phases of clinical trials and 5 approved drugs. For 35 discontinued drugs, we could establish links to 121 clinical trials in ClinicalTrials.gov (in March 2016). For the remaining 24 discontinued drugs, we could not find any trial information with AD as condition (Supplementary information S3 (table)). However, even for some discontinued drugs where we could establish a link to clinical trials, essential information such as phase number and enrollment were missing. It is likely that many clinical trials have not been registered in ClinicalTrials.gov, and therefore we had to make assumptions about average trial sizes and associated cost for the estimation of the R&D investments. For example, clioquinol (also known as iodochlorhydroxyquin and PBT-1), developed by Prana Biotechnology (http://www.alzforum.org/therapeutics/clioquinol) underwent a phase II trial in 2003, although it is not registered in ClinicalTrials.gov. We have also undertaken substantial effort to fill in missing information: we have manually searched in relevant websites and publications for data such as number of patients enrolled or the phase of the clinical trials linked with the study. If information on enrolment numbers could not be found (not even with manual search), we assigned an average number of patients for each phase (Table 2). Assumptions about average trial sizes were made based on the range of empirically determined, prototypical AD trials.

In total, we were able to identify sufficient information on 37 discontinued drugs to use for analysis which provided the basis for the information shown in Figure 1a, and on 61 of the 88 drugs in ongoing clinical trials for the information shown in Figure 1b.

In format provided by Hofmann *et al.* (doi:10.1038/nrd.2017.169)

Table 2 | Average number of patients enrolled in trials per phase

Phase	Number of recruited patients (empirical ranges of numbers)	Average number of patients we considered if information was not available
Phase I	20-100	60
Phase II	100-300	200
Phase III	1,000-2,000	1,500
Phase 0	10	8

Calculation of cost

After collecting and aggregating as much relevant information as possible, we calculated the average cost of a clinical trial for the drugs by multiplying the number of enrolled patients with the average cost per patient.

References

1. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9**, 203–214 (2010).

2. Adams, C.P. & Brantner V. V. Estimating the cost of new drug development: is it really 802 million dollars? *Health Aff.* **25**, 420–428 (2006)

Summary

Despite the fact that drug discovery and development in AD is characterized by low success rates, we also observe a very limited spectrum of disease-related, "core pathophysiology" processes currently being prioritized. The fundamental question, I was investigating to answer in the previous chapter is: are we targeting the right mechanisms or do we need to broaden the strategies in drug discovery in order to invest more on novel mechanisms rather than the established ones which have to date been unsuccessful? The intention behind this work is to make the levels of pharma and biotech R&D investments in putative disease mechanisms underlying neurodegenerative diseases more transparent. I am able to generate this different view on "targeted disease mechanisms", because we have generated a large "cause-and-effect" model of AD in the past as depicted in chapter 1 and 2. Annotating the subgraphs of this model with R&D spending provides a basis for a "mechanism-centric" discussion of drug discovery and development strategies in AD. Of course, we intend to stir up discussions about "where the money goes". The approach we took is not trivial: the calculation of the cost for drug development is challenging and a lot of information is missing. Different publications communicate rather diverse estimates for the overall costs of drug discovery and drug development. Neurological diseases seem to be particularly challenging, when it comes to cost estimates for drug development.

An interesting aspect of our results is the option, to specifically look into mechanisms that are "under-valued" or at least "not targeted". We believe that the systematic analysis of "shared mechanisms" between AD and other medical conditions may open promising new routes to "re-purpose" drugs and to "re-use" therapies already being tested in clinical trials for other conditions.

Of course, we are aware of the concerns and limitations (e.g. trial design; possible clinical readouts; legal aspects of changes in the approved trial protocol), but we believe that the pressing need to come up with a true "cure" or at least a "true prevention" for AD justifies the effort.

Conclusion and Outlook

Alzheimer's disease (AD) is a multi-factorial, complex pathological condition, which primarily affects an elderly population (Alz.org, 2017). Nevertheless, more than 5% of Alzheimer's cases in the US, have an early onset starting at an age ranging from the 40s to 50s (Alz.org, 2017), indicating that genetic predispositions contribute to a significant degree to the incidence of AD.

Vast amounts of data and knowledge on AD are available as a result of high-throughput omics studies and the overwhelming amounts of research efforts being conducted in the AD field. The task of deciphering the underlying mechanisms and interactions among various factors from this huge volume of data and knowledge becomes a nearly impossible challenge for an unaided human. Therefore, we need an integrative platform which reduces the dimension of data and knowledge, without any loss in expressiveness. Bioinformatics modelling approaches provide us with this integrative platform, however, the expressiveness of the relationships that occur in a biological context are often not comparable. The conventional modelling approaches merely provided associations and interactions among entities, but no cause-and-effect relationships. In this work, I have used a computational modelling approach known as Biological Expression Language (BEL) to model the knowledge in AD in a highly granular fashion, integrating knowledge from various biological levels. BEL is a knowledge integration language that uses the power of computing to make vast amounts of knowledge more approachable to humans.

The AD model built during the course of this work is the first of its kind and most likely the largest computable AD model in the world as of now. In addition to the AD model, I have generated and published a normal physiological model of the human brain. This was used to

compare the pathological changes which occur in AD as well as to portray the research bias that exists in the AD research field. Using these models, I have identified an early perturbed mechanism of AD as well as the functional role of genetic and epigenetic variants in a disease context. Apart from the two mechanistic hypotheses, which are published, this project subsequently has delivered more than 120 candidate mechanisms potentially involved in the aetiology of AD and is now made freely available to the research community through NeuroMMSig, a server for mechanism enrichment in clinical data.

These models are valuable contributions to the AD research community as the scientific knowledge included is reusable, can be further extended, can be converted into different formats and can serve analyses purposes using different algorithms. The model I developed served as one of the key cornerstones for the IMI funded European project; AETIONOMY [https://www.aetionomy.eu]. Through this model the identification of specific mechanisms dysregulated in AD was made possible. The computable cause-and-effect modeling approach described in this work can be complemented with a systems biology simulation at the systems level. The construction and simulation of computable cause-and-effect models of disease pathology may therefore be able to predict the response to a drug or effectively contribute to the design of clinical experiments.

Due to the limited access of human brain tissues and the immense complexity of the brain, mouse models have been extensively used as preclinical models in AD clinical trials. However, taking into consideration repeated drug failures, I have – in a well-defined, highly interesting pathophysiology context - compared systematically the mouse and human brain at molecular, cellular, and pathway levels to assess the functional equivalence between the two species. My findings imply that although the two species share some similarities at the cellular or pathway

level, the pattern of molecular interactions that form, govern, and regulate those pathways are substantially different between mouse and human. It is important to note that our analysis is purely based on the research and publication bias in mouse and human experiments. However, the in-silico models as developed in this work can effectively contribute to a well-informed design of *in vivo* mouse models that still needs to be shown to be predictive for human disease conditions.

Although we have computable or preclinical models available to predict correct drug targets, we also observe a very limited spectrum of disease-related, "core pathophysiology" processes currently being prioritized as drug targets. During the course of this work, I have mapped the landscape of 126 mechanisms comprising the AD computable model, targets and R&D investments and showed how the AD research community has been hesitant to move to new and emerging biology/targets. This work established how wide the spectrum of possible candidate mechanisms to be targeted could potentially be. It is therefore a fitting time to analyze the spectrum of opportunities that exists for the targeting of new, disease-associated mechanisms in AD.

Taken together, my thesis work presents a comprehensive computable model for AD research, which comprises of more than 120 different mechanisms of AD of which most could be targeted. These mechanisms included various scales of biology starting from genomic levels to phenotypic levels.

However, validation of the established mechanisms or hypotheses in wet labs still remains to be done. This works lacks the systematic comparison of the computable model of AD to other existing pathway databases, such as KEGG or REACTOME. In future, it is also needed to connect the mechanistic model with existing databases to do more advanced analysis and ensure

the completeness of the model. More research needs to be invested in these areas to bridge the gap between the molecular and clinical level data to the mechanistic information encoded in these models. For example, the model needs to be enriched more with clinical indices to connect with data like the Alzheimer's Disease Neuroimaging Initiative (ADNI). Another prospect for the future based on this work is to provide a longitudinal aspect, where each relationship in the model could be assigned to a certain stage or time point based on the course of AD. If we are successful in linking clinical data with the mechanistic aspects, then a simulated trajectory showing the progress of a patient along with his age is not a distant. Lastly, there is an immense need to continue to update the knowledge model automatically as new insights are gathered in the research field.

Reference

Alz.org, 2017 :*Alz.org.* Zugriff am 5. October 2017.
https://www.alz.org/alzheimers_disease_what_is_alzheimers.asp
https://www.alz.org/alzheimers_disease_early_onset.asp